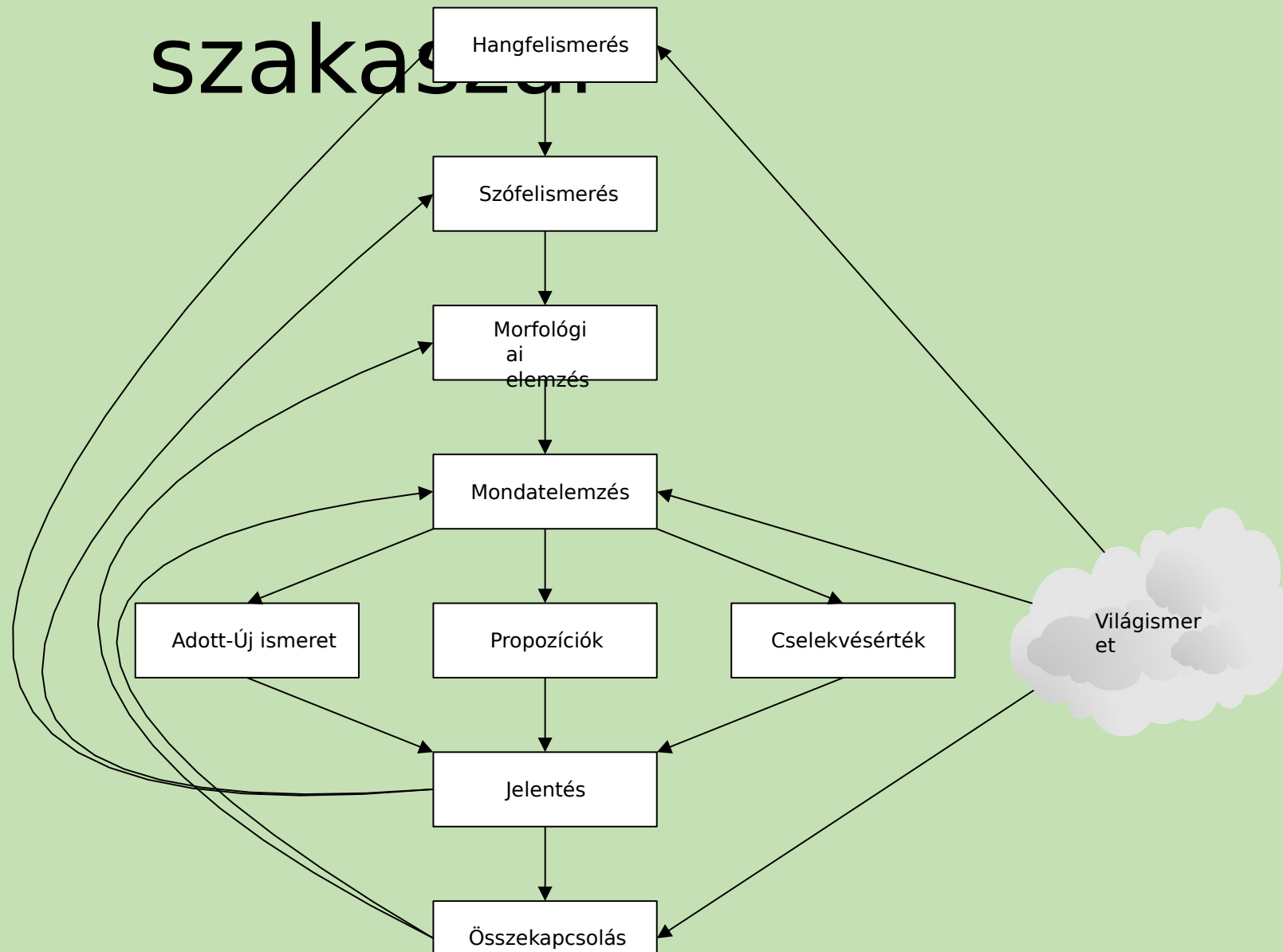


A nyelvi
megértés
folyamata:
információ és
szabályok
Pléh Csaba

vispleh@ceu.edu

- Előadás a Paletta sorozatban.
Wesley János
Főiskola
- 2020.Október 17

A megértés szakaszai



Statisztika és szabályok örök versenye

- RENDEZETLENSÉG
CSÖKKENTÉS

- A szófelismerés információs értelmezése.
- A morfológia információelméleti értelmezése. A morfológiai modellek kvantitatív értelmezése.
- A mondaton információs megközelítése.
- A priming folyamatok általános érdekessége.. Szövegmegértési sémamodellek és az információ. Kintsch információintegrálási felfogása.
- Szabály és statisztika kettős modellek. Pinker, Ullman.

- A VÁRATLAN KEZELÉSE

- A megértés folyamatának komponensei és a nyelvi szintek
- A megértés gyorsasága. Találatok és Tévedések versengése
- Alternatív felfogások: modulok és interakciók
- A tudás és az inger viszonya. Elvárások a megértésben
- Lexikai emlékezet, világismeret és információ

Eliza:
„pszichoterape
uta”
Weisenbaum
1966. mlt
hallunk bele
statisztika
alapján?

A férfiak mind egyformák.

Ezt hogy érted?

Mindig nyaggatnak minket valamiért.

Mondanál egy konkrét példát?

Hát, a fiúm küldött ide.



A fiúd küldött ide.

Azt mondja, mindig depressziós
vagyok.

Sajnálom, hogy depressziós vagy.

Talán ki kellene békülnöm az
anyámmal.

Mesélj még a családról.

- 
- A módszer:
 - Kulcsszókeresés
 - mindig
 - anyám
 - Kulcsszó aktivál egy szabályt
 - Mondanál egy konkrét példát?
 - Mesélj még a családról.
 - Reguláris kifejezések segítségével behelyettesítés a szabálynak megfelelően
 - fiúm -> fiúd
 - A kulcsszavak rangsorolva vannak
- 

Mi ELISSA tanulság a?

Sok emberi folyamat algoritmizálható

Nem is olyan elvont az ember

Kérdés, hogy olyan-e ez, mintha átmenne a Turing próbán?

Mi a statisztika szerepe az önbecsapásban?

Hogyan lehet becsapni?

És a valódi pszichológust?

Vissza a laborbaa

Találatok és Tévedések

- A kontextus
versengése
támogat : Találat
- A kontextus
félrevezet
- Tévedé
s

kutya

- Kinyitom ajtót a
kuxxxx



kulccs
al



kutyán
ak

A megértés folyamatának komponensei. A nyelvi felfogásban

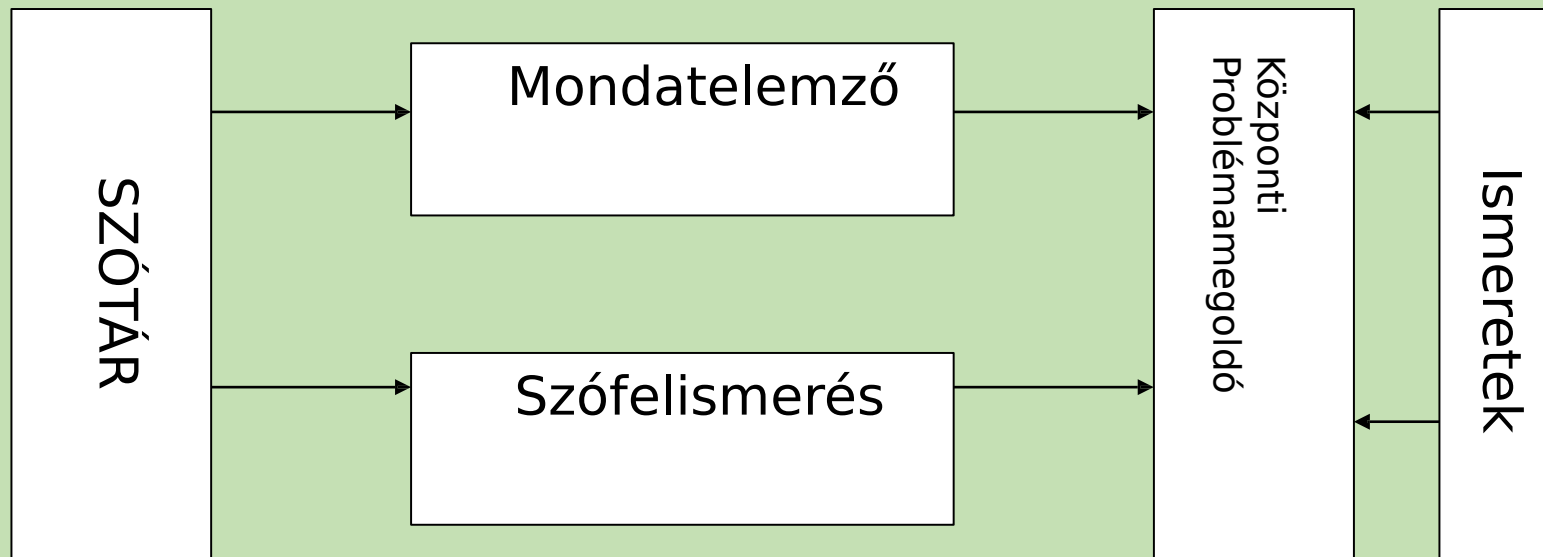
- **Moduláris felfogás**

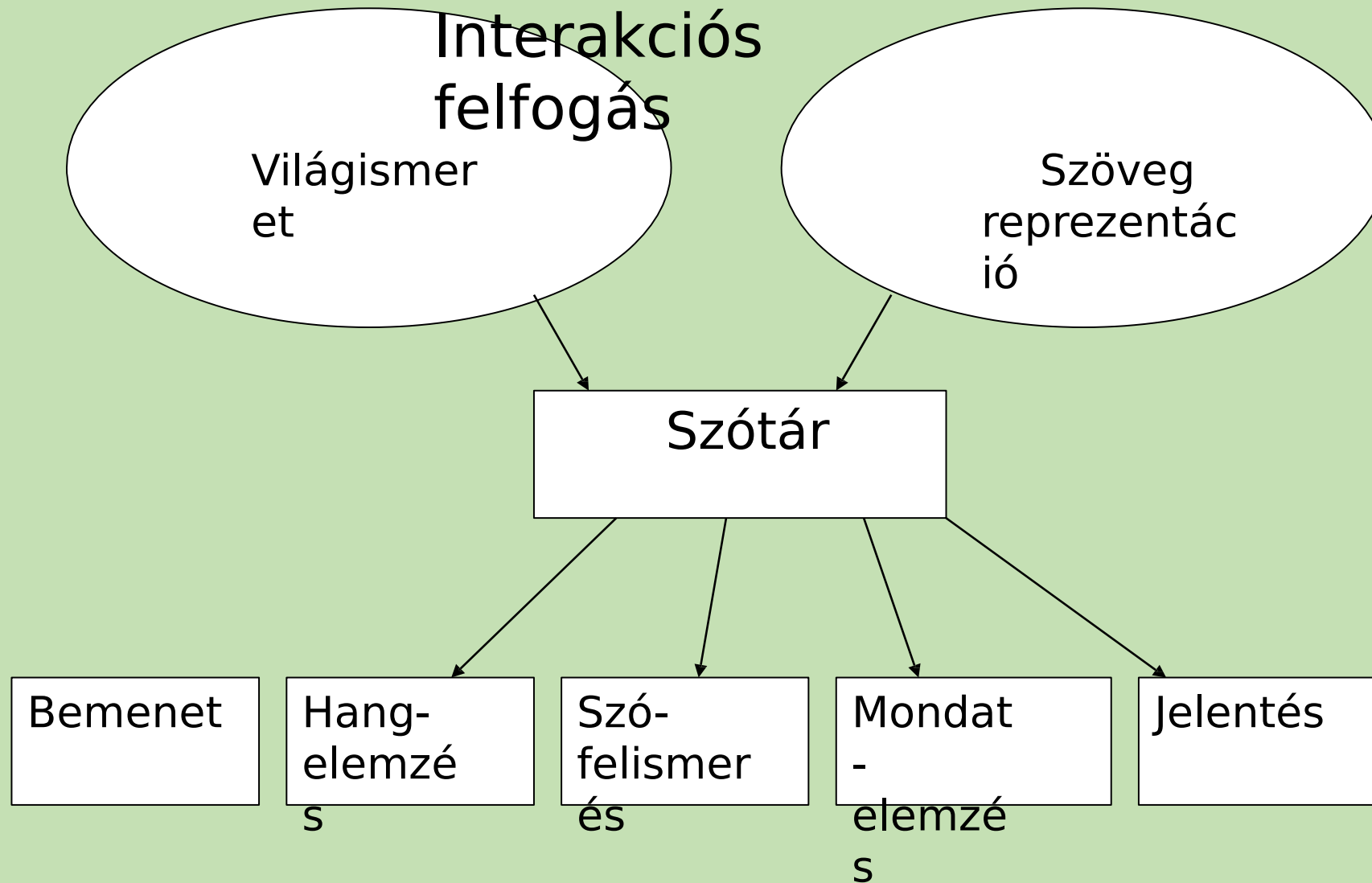
- Rögzített összetevők
- Alulról-felfelé
- Forma-jelentés
- Ostoba feldolgozás
- Algoritmikus összetevők

- **Interakciós felfogás**

- Hajlékony összetevők
- Felülről-lefelé is
- Elvárások döntőek
- Értelmes integráció
- Heurisztikus megoldások

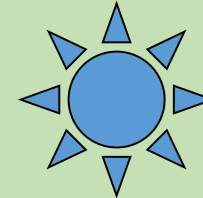
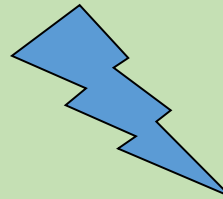
Modularista modell





A nyelvészeti információelméleti érdeklődés szakaszai

- **Korai lelkesedés: Nyelvészeti kritika**
Visszajön a statisztika
a statisztika



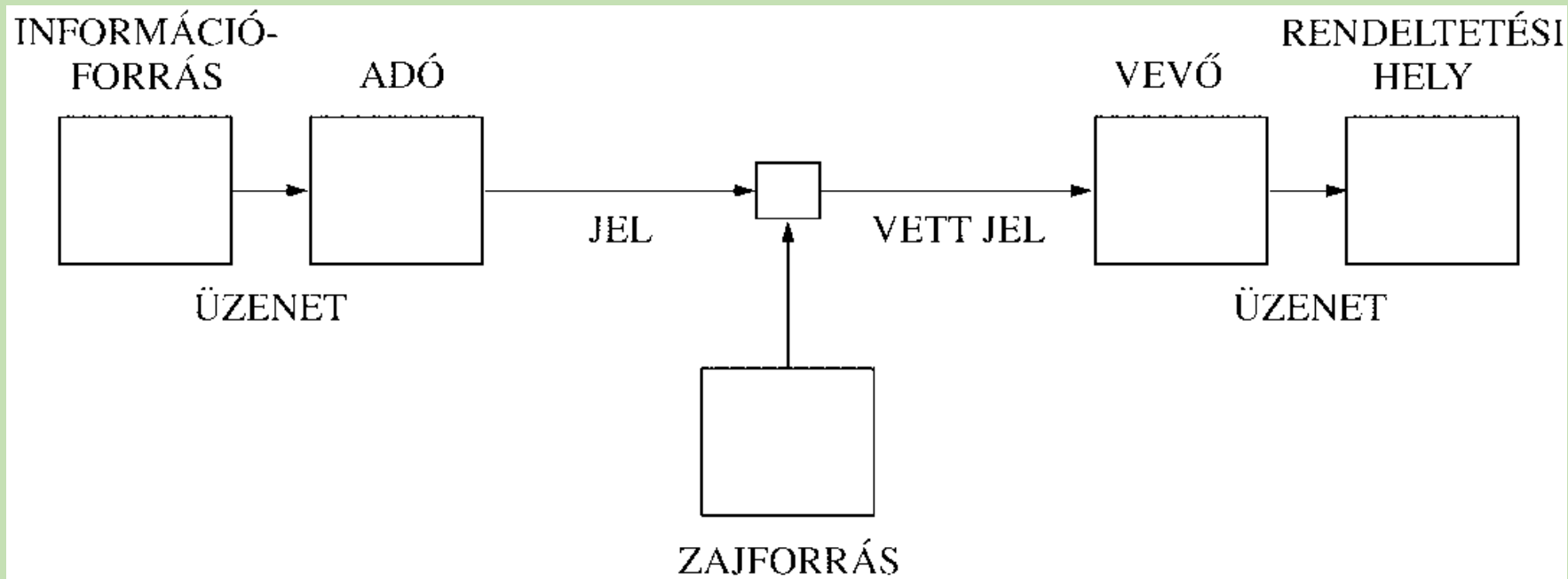
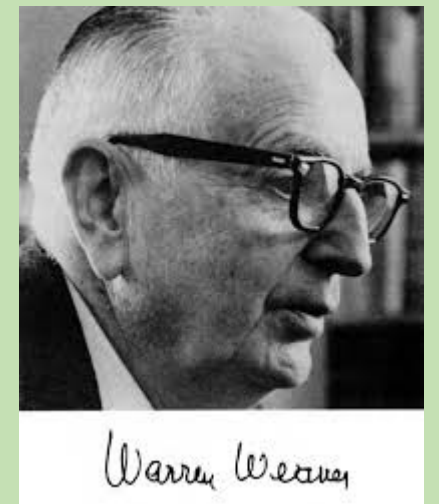
•1950 Shannon
• Miller

1960 Chomsky
G. Miller

1990 Kostic
Bayern Saffran

Kiinduló Shannon Weaver modell

? Hol a kód?
? Hol a jelentés ?



Alapfogalmak

- Információ és bizonytalanság
- Rendszerek átlagos információértéke és a lehetséges jelszám. ÁBC nagyságtól függ
- Átlagos információ $\log N$
- Bináris egységekben, kettes alapú log.
- Miért?

neuro

optimális keresés

természeti folyamatok csillapítása

Találjuk ki

Gondoltam egy 0... 9 közti számot

Gondoltam egy betűt

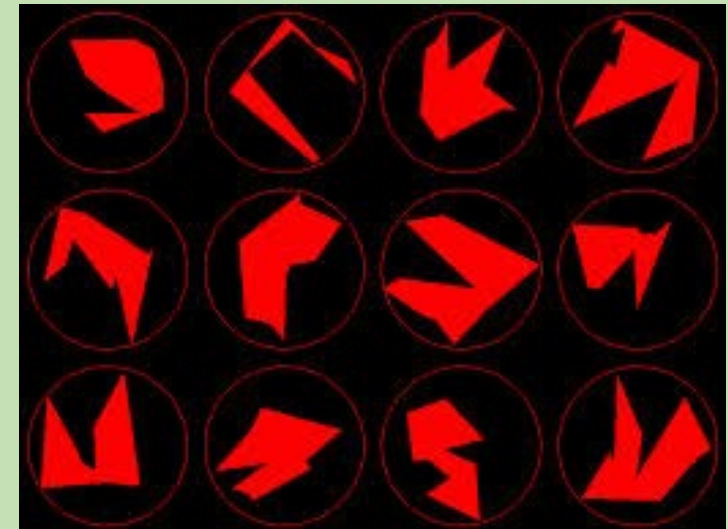
képlete

Az entrópia általános

$$H = - \sum p_i \log p_i$$

Entrópia változásai

- Maximális ha minden p egyforma
- Ha vannak kiemelt valószínűségű jelek, csökken a rendezetlenség
- Redundancia: a $H(\max)$ és a $H(\text{megfigyelt})$ viszonya
- A struktúra kérdése vizuális formáknál Attneave és Garner
- Akkor a Gestalt szerveződés alacsony információ lenne?
- Vagy éppen a hirtelen
- változás adja a formát?

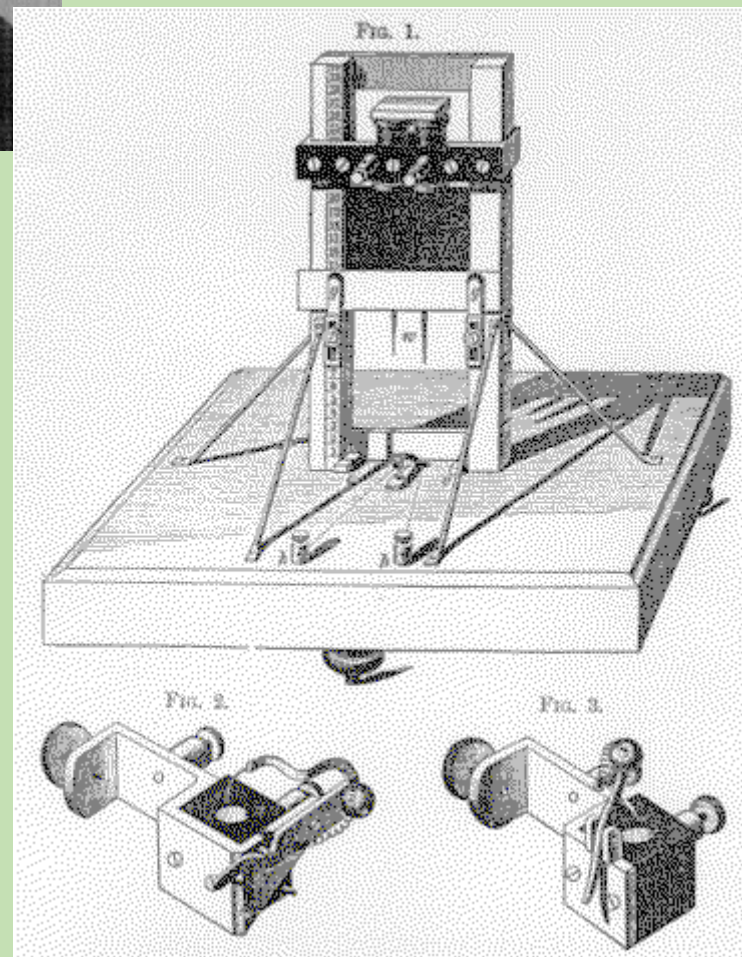


A statisztikai
múlt az
észlelésre : A
gyakoriság
szerepe

- Cattell, 1886:.. Szavak olvasása. Még embert vizsgálunk, de már a rendszer is érdekel
- *It will be noticed that the perception-time is only slightly longer for a word than for a single letter; we do not therefore perceive separately the letters of which a word is composed, but the word as a whole. The application of this to teaching children to read is evident*

TABLE XXVI.

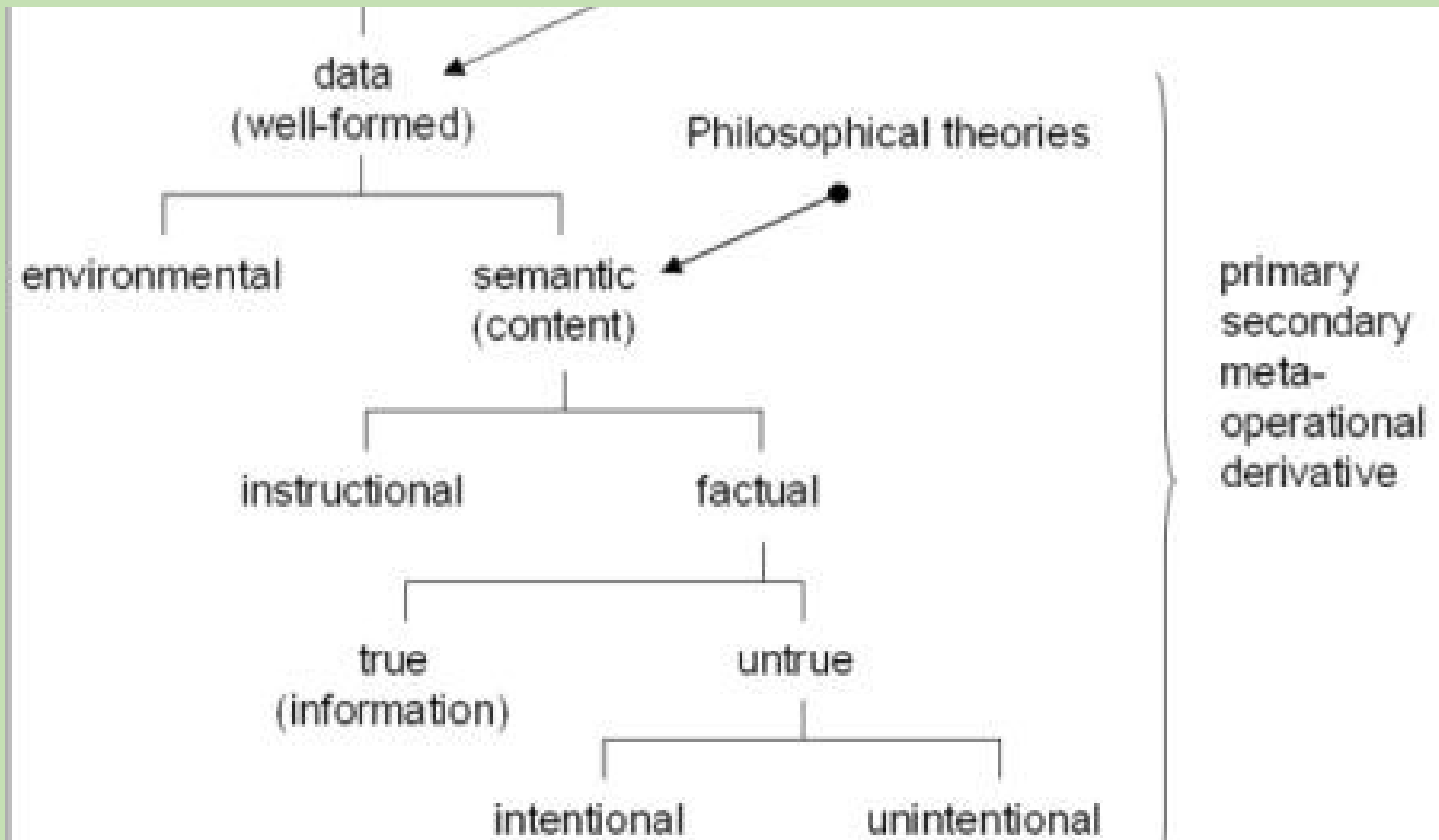
		HAND.					LIP-KEY.					
			B		C			B		C		
			R	R'	R			R'	R	R'	R	R'
Mind.....	12. XII.	353	352	337	329	13. I.....	360	366	374	364		
Life.....	15.....	348	351	373	377		366	367	363	365		
Time.....	16.....	333	330	375	372	15.....	311	312	371	366		
House.....		377	366	383	389		331	324	355	361		
Child.....		345	343	328	339	17.....	347	341	370	375		
Year.....	18.....	353	359	369	360		337	336	354	358		
Truth.....		352	329	376	367	29.....	302	311	360	353		
Name.....		341	339	392	393		313	315	374	380		
Light.....	19.....	332	328	327	323		325	332	372	372		
Ship.....		318	313	336	332		294	302	340	340		
A.....		345	341	360	358		329	331	363	363		
AV.....		24	13	26	17		23	12	28	20		
False.....		2		4			7		0			
Education....	5. I.....	331	331	346	348	17. I.....	349	345	382	386		
Philosophy...		330	322	349	354		347	351	376	377		
Knowledge...	7.....	341	337	366	360	22.....	353	348	329	319		
Architecture.		377	375	382	377		357	355	336	340		
Literature...	10.....	339	320	363	354	23.....	333	332	377	382		
Temperance.		341	333	399	404		339	330	377	376		
Ignorance....		300	297	380	369		325	319	378	382		
Physician....		325	329	380	375	26.....	339	333	351	346		
Enthusiasm..	12.....	334	337	405	409		353	349	409	400		
Imagination.		321	317	384	375		342	337	395	391		
A.....		334	330	375	373		344	340	371	370		
AV.....		25	16	28	19		23	15	27	17		
False.....		8		8			6		9			
Buch.....	24. I.....	290	294	367	363	23. I.....	315	318	359	355		
Zahl.....		309	311	380	378		310	319	370	378		
Kunst.....		307	309	369	374		310	314	362	352		
Welt.....		308	307	361	353		308	305	362	362		
Haus.....	26.....	295	292	354	353	24.....	299	297	339	344		
Licht.....		324	323	354	359		330	329	356	350		
Kind.....		323	323	377	380		303	308	352	356		
Land.....	29.....	309	307	363	365	26.....	316	321	373	365		
Traum.....		321	316	377	376		324	325	368	373		
Jahr.....		319	318	365	368		321	325	374	378		
A.....		311	310	367	367		314	316	362	361		
AV.....		14	9	20	13		17	12	31	20		
False.....		6		5			10		7			



A szavak gyakorisága gyakran felmerül

- Olvasás és szókincs tudás mérés megalapozásához
- Asszociatív adatok értelmezéséhez
- Betegek bizarr beszédének interpretálásához
- Még stílus statisztikák is megjelennek
- Legfeljebb szófajban jellemezve a szavak

Az információ sokrétű fogalom



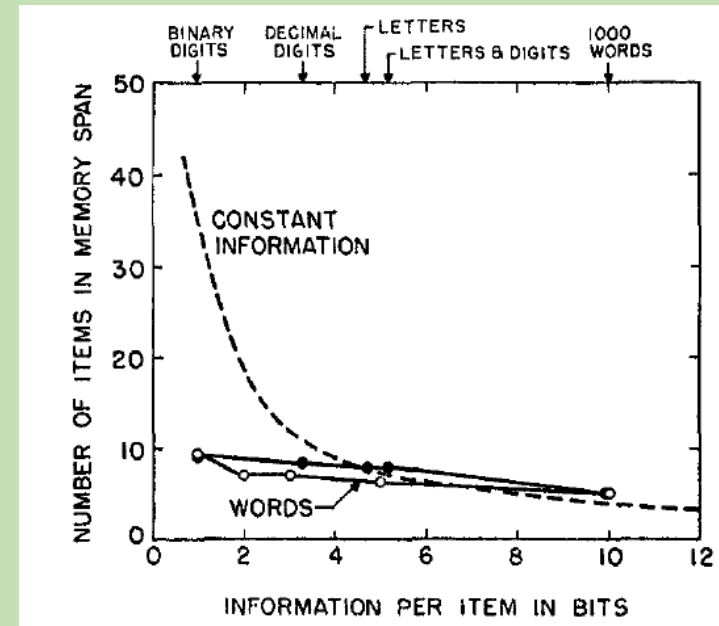
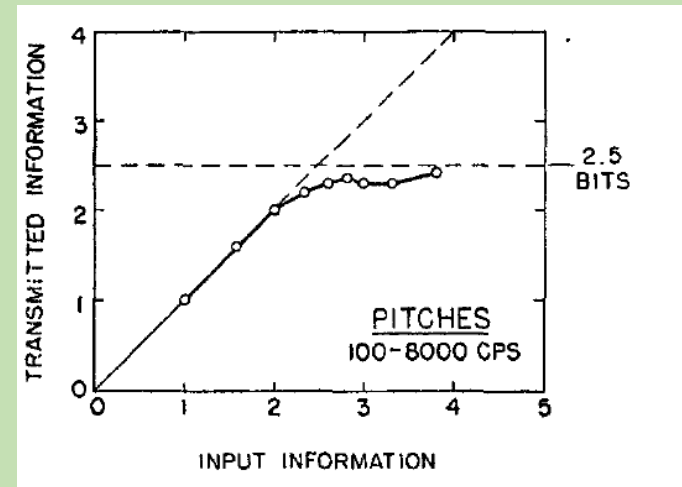
- Miért ,szintaktikai' a Shannon féle?
- Mi informatívabb
köznapiilag: kutya
vagy állat? S melyik
ritkább?
- Szemantikai elméletek
- Bar Hillel MacKay
- Igazság és p ill.
diszkriminatív érték

Mire vagyunk limitálva? Miller felismerései

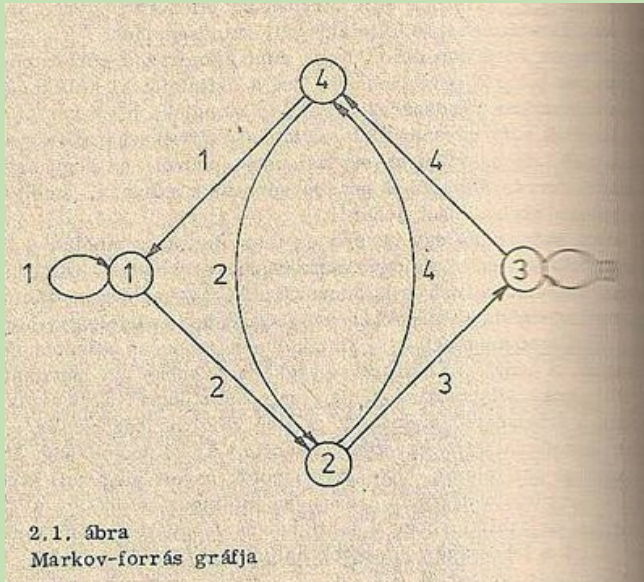
- Abszolút ítélet: 2-4 bit, 4-8 szint



- Munkamemória: 7 ± 2 egység, s nem info csatorna. „Tömb” határ



Miller és Chomsky megfordítják a trendet: mindez legfeljebb egy gyenge Markov modell



- A Chomsky hierarchia és a pszicholingvisztika
- Nemcsak a nyelvtanra, hanem a nyelvi viselkedésre is algebrai modell kell
- Abból is nem akármilyen
- Véges állapotú nem elég
- Az ellenérdekeltek szerint igen, ha a gráf ágaihoz valószínűségeket rendelünk

Visszatérő kérdések minden elemzési

1. Az általánosítás szintje, hangtípus, szófaj
stb

2. Statisztika és szabályok viszonya

3. Alulról-felfelé és Felülről-lefelé
áramlás

4. Helyzethez igazodás és tanulás

A kontextus hatások a hangok, szavak, és

szerkezetek szintjén is
Gyors a beszéd sebessége a megértés. Percenként 140-180 szó, 3-4 szó/sec. Mesterségesen felgyorsított beszéd: akár a normális sebesség kétszeresével mondott beszédet is megértjük.

Pollack és Pickett (1964) természetes beszélgetésből kivágott egyedi szavakat izoláltan mutattak be: felismerhetetlenek voltak. Mondatba visszarakva tisztán felismerhetőek voltak.

Nyelvtani szerkezetek és a statisztika teszi lehetővé, hogy jóslásokat tegyünk beszéd közben. A leggyakrabban használt 50 angol szó a beszélt szövegek 60%-át, az írottak 45%-át teszi ki.

A kulccsal kinyitottam ⇒ STATISZTIK

Ajtó
A kulccsal kinyitottam ⇒ az ajtót

NYELVTAN

Tévedések a szófelismerésben Gósy Mária gyűjtése

szint	eredeti forma	meghallott forma	értelmezés
hangészlelés	gőzös hajó	közös hajó	hang hasonlóság
szófelismerés	vélelem	védelem	gyakoribb szó
morfológia	fejez	fejes	gyakoribb rag
morfológia	leszedték a záradakat	leszedték az áradakat	hibás tagolás
mondatelemzés	kerget a lányod	kergeti a lányodat	egyértelműsítés
jelentés	hallod lám	hallomás	értelmezés
szövegösszefüggés	csak egy játékszer	csak egy pár ékszer	ékszerboltban

Statisztika és szabályok viszonya

A hangok elvárt sorrendjei:

- Ritka *kra pro*
- Vajon ez segíti vagy nehezíti-e az észlelést?
- Statisztikai tény-e, hogy az – *ik* és az – *ás* szóvég gyakori, vagy nyelvtani, szabályalapú hatás?

* kp , tv

Áttekintés az entrópia kihívásról

- Kezdeti lelkesedés. Chomsky kritikája
- Az utóbbi évtized a nagy digitális szöveganyagok révén: Shannon reneszánsz
- Négy példa:
 - Szófelismerés Akusztikus bejósági (gating) és olvasott szavakra vonatkozó szótári döntés
 - Alaktani döntések és entrópia
 - Szótári és alaktani entrópia viszonya
 - Vonzat bejósítás és bizonytalanság
- Szerkezeti megfontolások mellett az entrópia változás segítségével magyarázni.

Csaba Dani Dani Kornél Klára Judit István



Entrópia



Gating



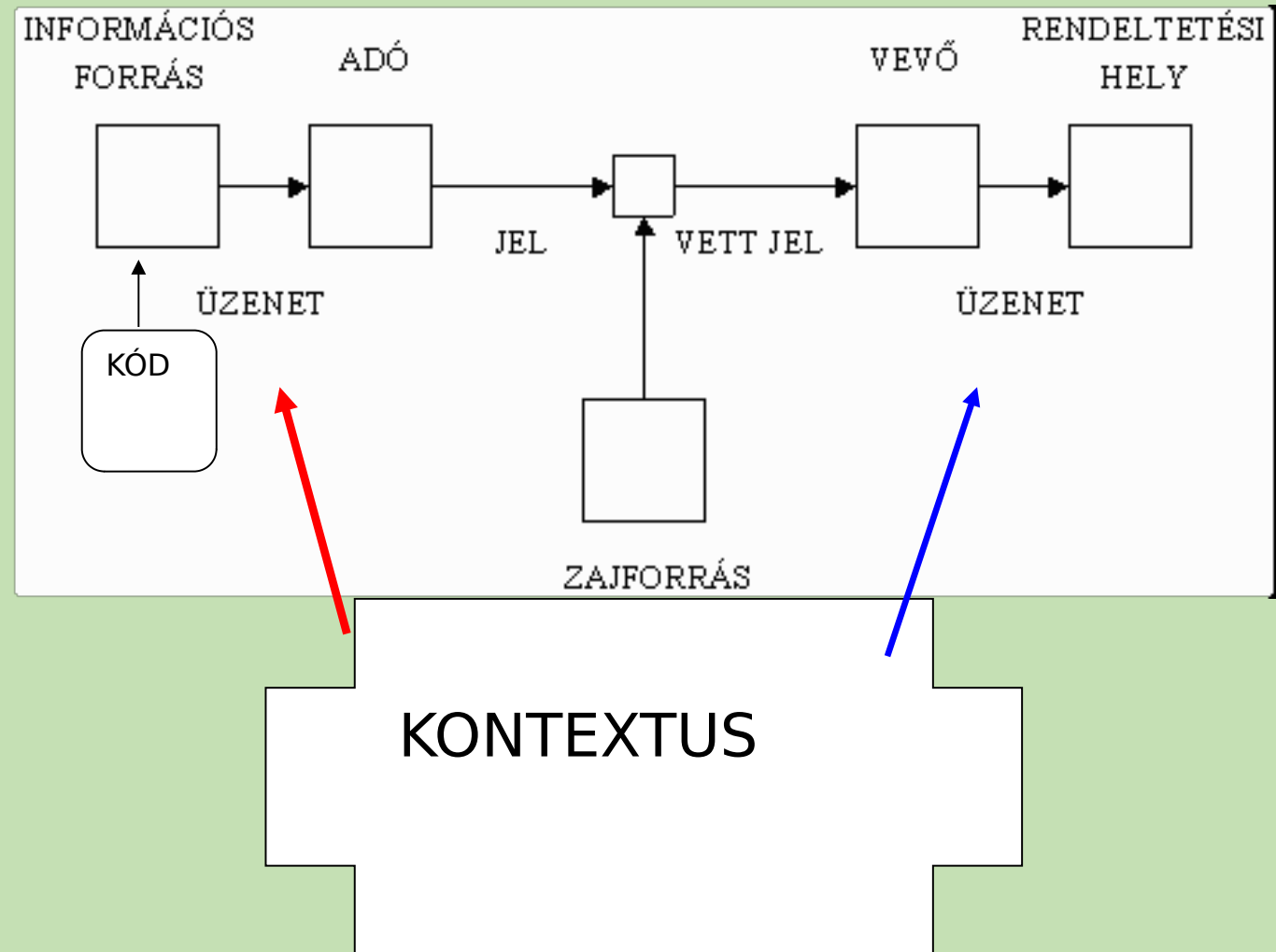
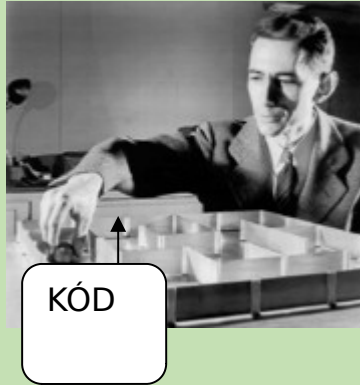
Priming



LexDec



Információelmélet és nyelv Shannon korában



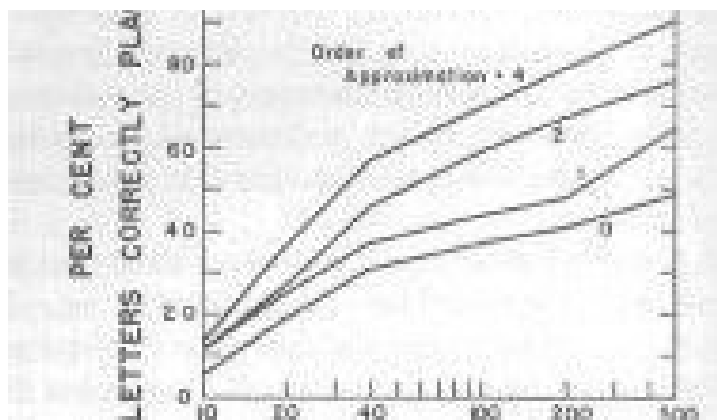
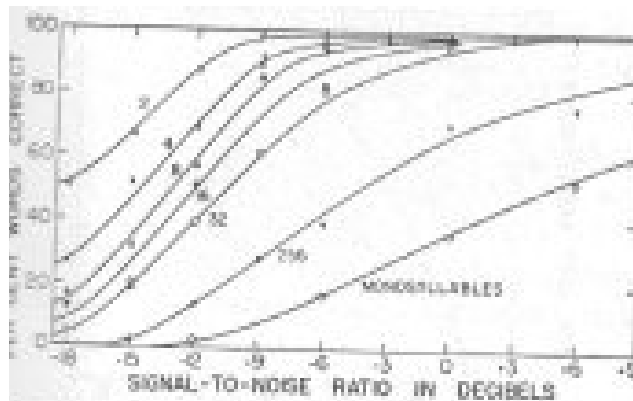
Nyelvi következmények

- 1. feldolgozási korlát és átvitt információ
- 2. redundancia és szófelismerés
- 3. statisztikai közelítések a nyelvhez és a feldolgozhatóság

Néhány
klasszikus
szófelismeré
si hatás
Miller, 1951
alapján

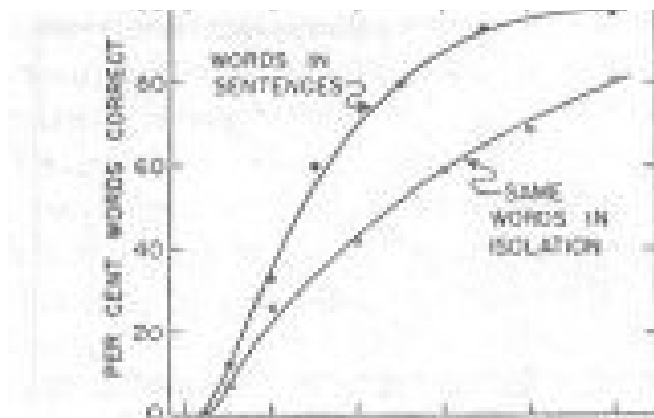
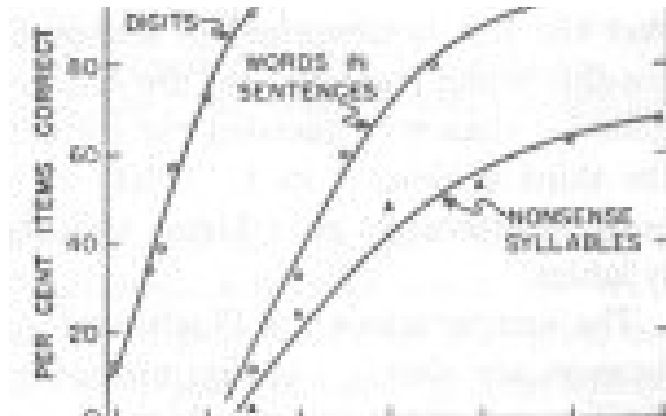
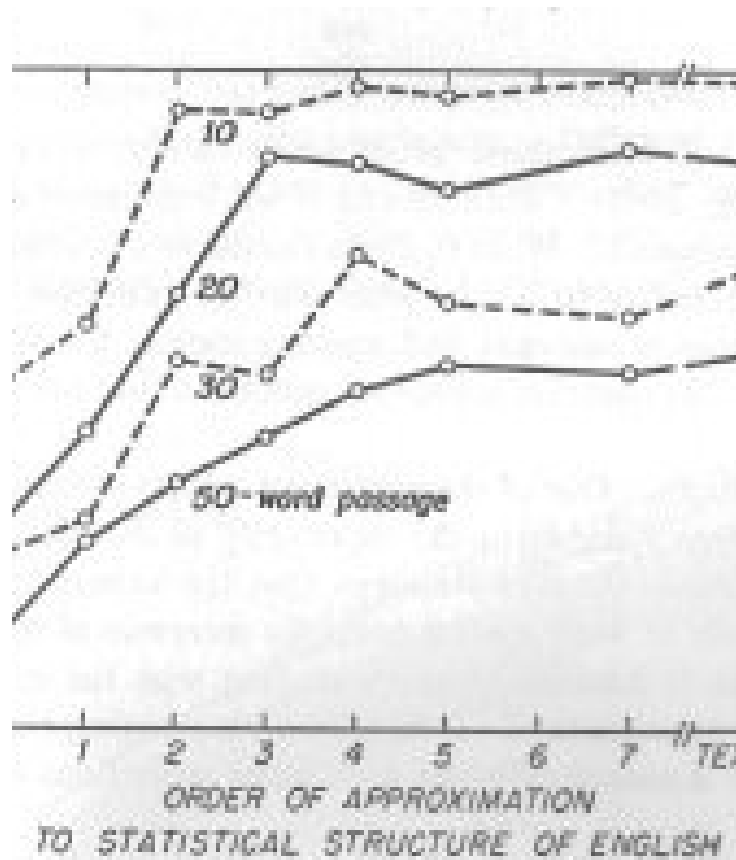
Hatás jellege	Modalitá s	Változók	Fő eredmény
Expozíciós idő a gyakoriság függvénye	Vizuális	Gyakoriság és felismerési idő	Logaritmikus függvény, megfelel az információnak
Zaj rontja a felismerést	Akuszt	Jel-zaj arány, szóhos sz	Felismerés zajban romlik
Expozíciós idő javítja a felismerést	Vizuális	Idő hatása 0-55 ms Statisztikai közelítés	Idő javít, de a statisztika hasonlóság ugyanígy
Nyelvtani szerveződés s javít	Akuszt kus	Jel-zaj arány, szótag, szavak, mondatok	Mondat >szó>szótag zajban
Mondatkontext us	Akuszt kus	Jel-zaj arány, elszigetelt szó és mondat	Mondat segíti felismerést

Információelméleti észlelés kutatás



- G. A. Miller
- Zaj és észlelés
- Expozíciós idő és felismerés

Kiterjesztve a nyelvre



- Értelmesség hatás
- Szavak és mondatok
- Statisztikai közelítés szintje

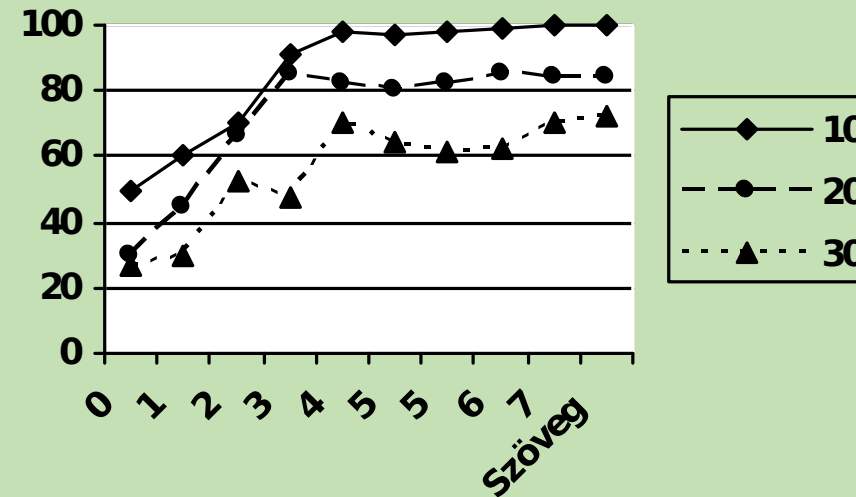
Statisztikai közelítések

0-d rendű: *mellékesen
következmény csinosan bankár
hajlik áram*

1. rendű : *képesség révén mely
engem keringő téged a varrás*

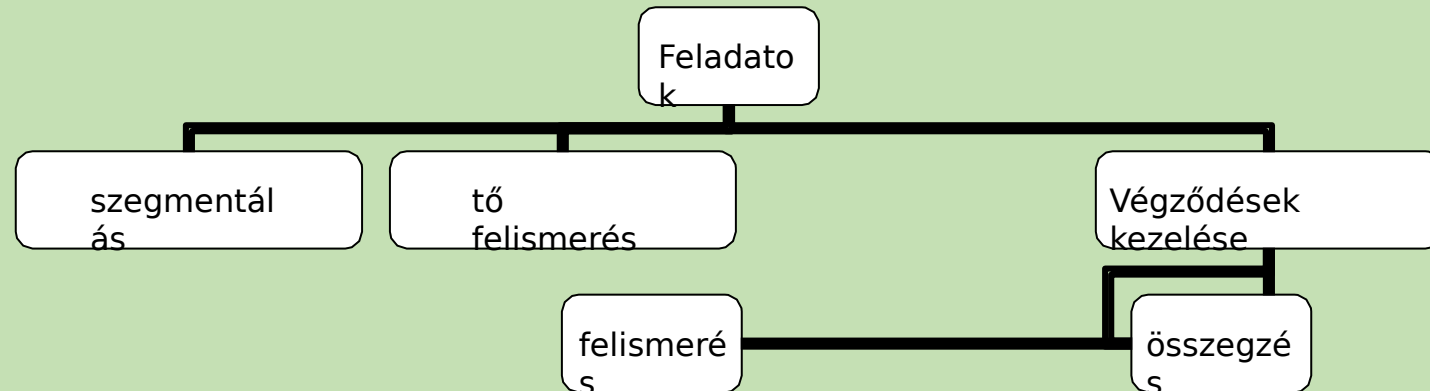
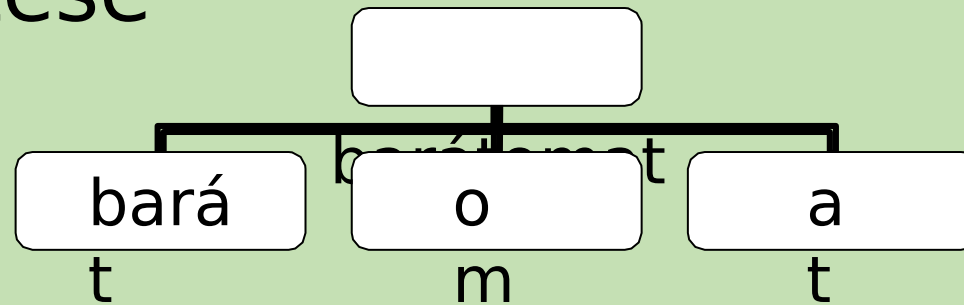
4. rendű : *láttam a meccset
befejezik szilveszter éjjel*

7 rendű : *ismerem zenei
képességeit miután bemutatta
előtte*




A szavak világa

- Tövek és végzódések: a szavak felépítése



Szavak felismerése és a mentális

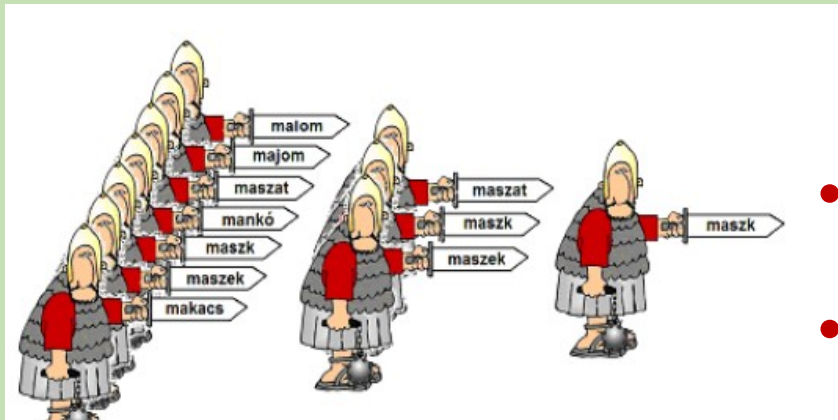
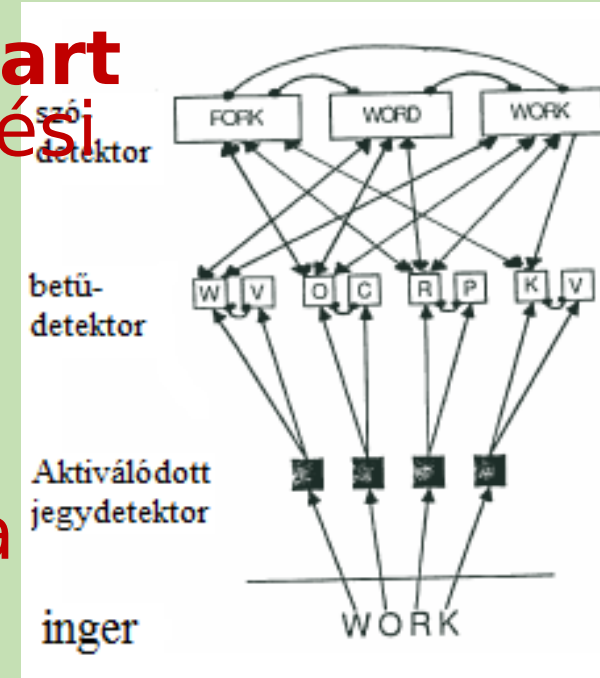
- **Alaptények:** gyakoriság, kontextus felülről-lefelé kutya  **kuvasz**
- Hozzáférási utak: akusztikus, grafémikus, közös.
- Általánosítás 1: a jelentés egy közös szótárban
Általánosítás 2: a *hó* és a *hav* a Hóhoz kapcsol
Általánosítás 3: a szófaj is tárolva van, s sok mindent elárul
pl. *kabát* *ik hisz a *kabát* főnév
- Küszöb és alapszint változások:
logogén kohorsz és

A szófelismerés modelljei

Cohort-modell - hallott szavak felismerése már az első szótag alapján

- **McClelland & Rummelhart (1981)** - vizuális felismerési modella versengésről

- **Marslen-Wilson (1987)**
 - Kompatibilis szavak automatikus aktivációja
 - Kiválasztás



- összefüggő szöveg - erős igazodás
- izolált szavaknál & értelmetlen szövegek - kevesebb az elvárási hatás

GATING - a paradigma

- Bevett szófelismerés vizsgáló eljárás
 - Nincsenek magyar adatok
 - Jól kontrollálható változók
 - **Francois Grosjean (1980)**
 - **melyek a szófelismerés kritikus pontjai?**
 - **személy szavakból hallgat egyre nagyobb részeket**
 - **Hatások:**
 - Gyakoriság (gyakoribbak – rosszabb teljesítmény)
 - Szóhosszúság (hosszabb szavak rosszabb teljesítmény)
 - Kontextus (mondatba foglalva jobb a teljesítmény)
 - **a növekvő, sorozatszerű, inkrementális bemutatásnak nincs hatása (Cotton és Grosjean, 1984)**
-

Ingerek

60 szó a
szószablyából

30 ritka (<152 ; $\sim 67,23$)

15 korai
egyediségi pontú

15 késői
egyediségi pontú

30 gyakori (>11982 ; ~ 55770)

15 korai
egyediségi pontú

15 késői
egyediségi pontú

szó	gyakoriság	Lemmaszám	gyed.pont	Első 4	entrópi
böllér	80	196	3	böll	0,603096
cécó	196	196	3	cécó	0
csöb	196	196	3	csöb	0
dunn	77	186	3	dunn	0,652352
dzsú	128	173	3	dzsú	1,139827
güzü	172	172	3	güzü	0,583207
gyűs	176	176	3	gyűs	1,472454
kégl	66	181	3	kégl	0
lucso	66	156	3	lucso	0
pócs	174	174	3	pócs	0
pócs	7	7	3	pócs	0
rücsö	8	45	3	rücs	0,4319
sass	128	128	3	sass	1,849419
üzér	14	34	3	üzér	0,443576
zsep	63	151	3	zsep	0
bögö	160	161	5	bögö	0
cinke	3	185	5	cink	1,948223
dublő	21	82	5	dublő	1,5944

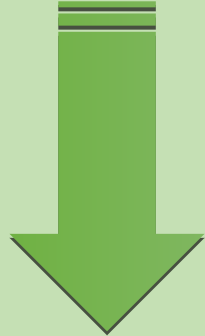
Ko
rai

szó	gyakoriság	Lemmaszám	gyed.pont	Első 4	entrópi
asszon	50211	17569	2	assz	1,540912
fórum	167634	167634	3	fóru	0
szoftv	31974	71251	3	szoft	0,076625
er	50381	141877	3	er	1,019915
pápa	25598	34252	3	pápa	1,069351
szféra	11602	26148	3	szfé	0,045589
tonna	11982	18232	3	tonn	0,597695
reg	28887	28887	3	reg	1,004682
keny	183	183280	3	keny	1,081985
típus	52944	52944	3	típu	0,009
műs	17263	17263	3	műs	0,837288
ablak	70943	70943	3	ablak	0,028
dollár	13876	38662	3	doll	0,008
ünne	12579	50454	2	ünne	0,00832
japán	17879	36339	3	japá	0,213908
család	71077	218999	5	csal	1,426039
kérdés	133771	573154	5	kérd	0,7521

A gyakorlatban

- **Kapuk**

- 90 ms
- 120 ms
- 210 ms
- 300 ms
- 390 ms



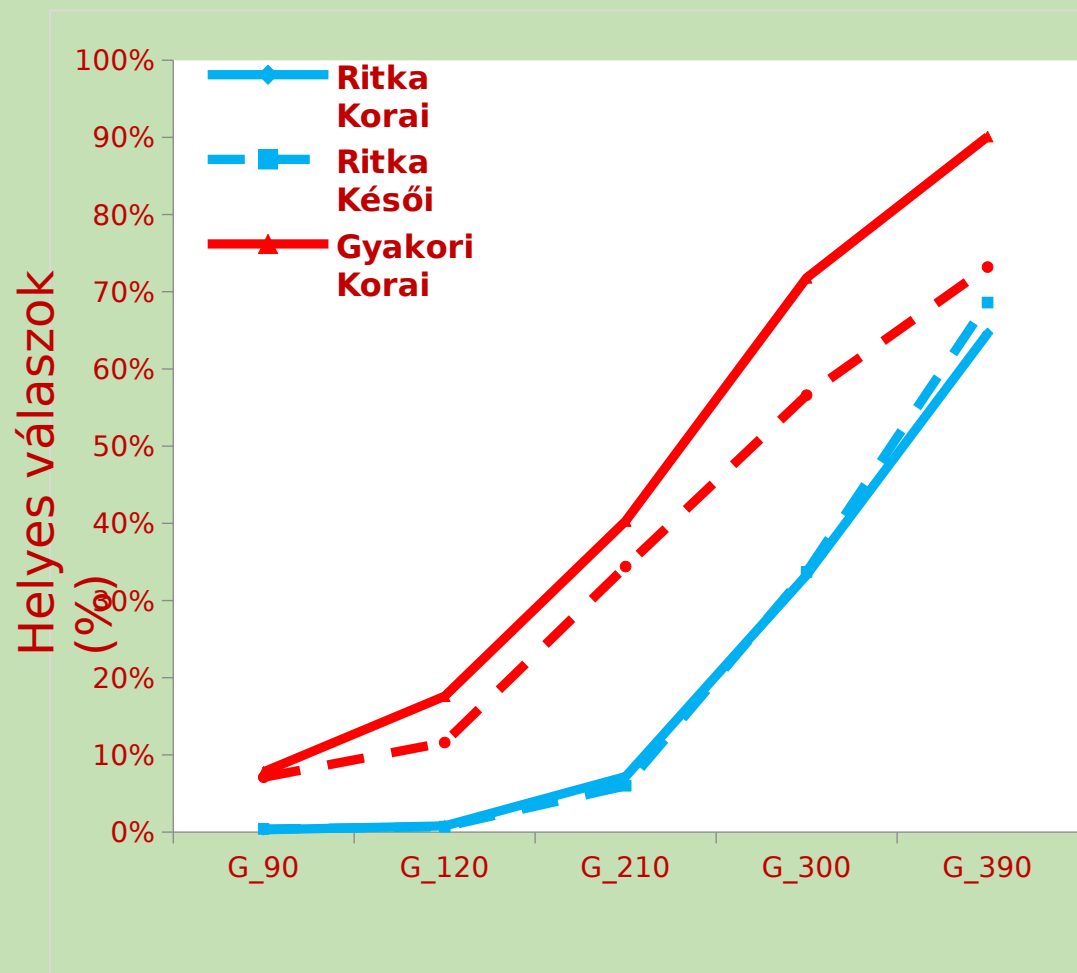
Inkrementális
bemutatás

- A szavak sorrendje random
 - Ha helyes volt egy adott kapunál a válasz (KV regisztrálja a **RI mérést követően (csak 2. kísérlet)** – SRB, mikrofon), akkor ugrik a következő szó 1. kapujára (90 ms)
 - **Konfidencia döntés**
 - 1 - egyáltalán nem biztos
 - 2 - inkább nem biztos
 - 3 - inkább biztos
 - 4 - teljesen biztos
-

Gating - a szögyakoriság és az egyediségi pont hatása a szófelismerésre

	90 (ms)	120	210	300	390	
• Gyakori – korai						ablak
• Gyakori – késői						nemzet
• Ritka – korai						böllér
• Ritka – késői						tartár

I. Kísérlet - pontosság

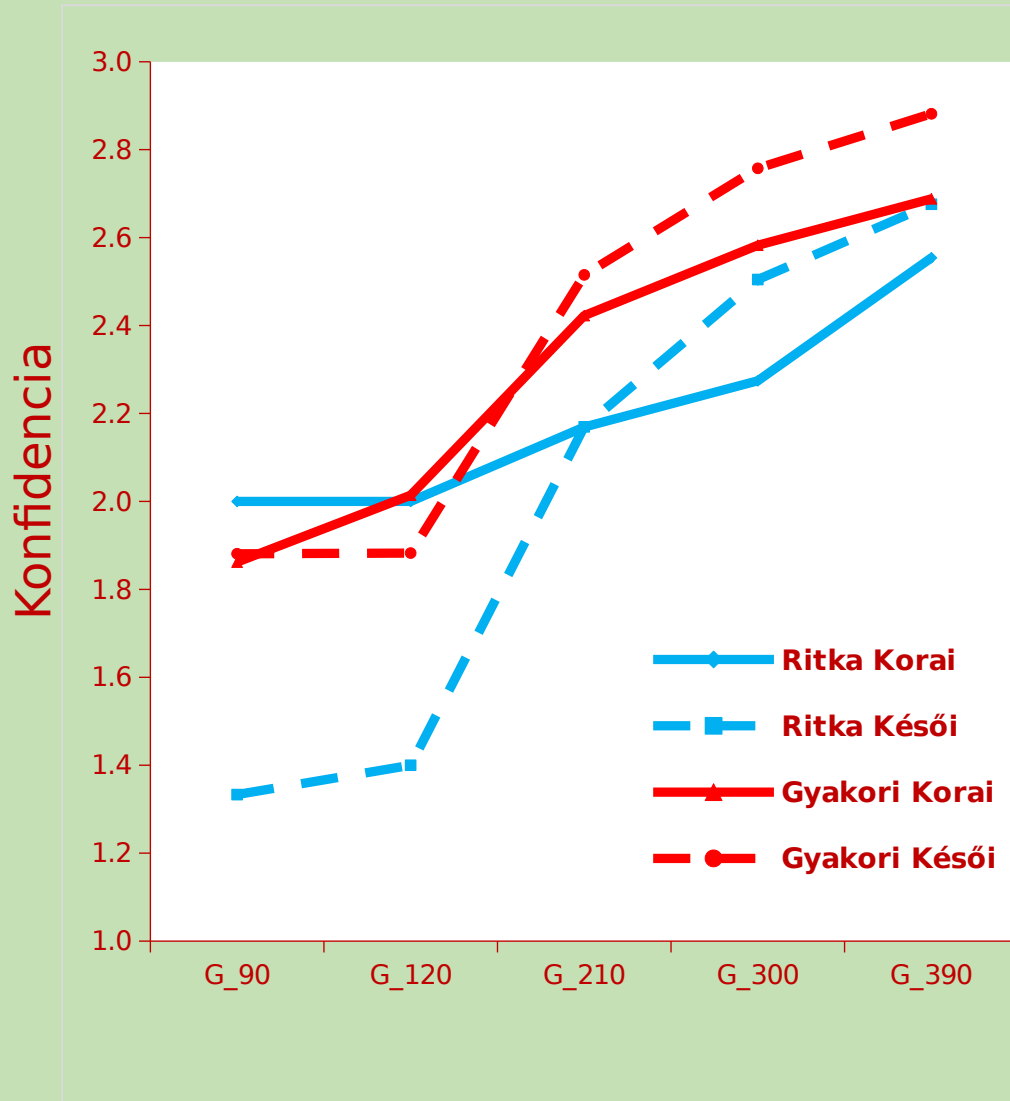


- Gyakoriság főhatás
- Egyediségi Pont főhatás
- Kapu főhatás
- Gyakoriság X Egyediségi Pont
- Gyakoriság X Kapu
- Egyediségi Pont X Kapu
- Gyakoriság X Egyediségi Pont X Kapu

$p < 0,001$

51 résztvevő; 20 férfi (21,12 év, $SD=1,37$) és **31 nő** (20,41 év,

I. Kísérlet - konfidencia



- Gyakoriság főhatás
- Egyediségi Pont főhatás
- Kapu főhatás
- Gyakoriság X Egyediségi Pont
- Gyakoriság X Kapu
- Egyediségi Pont X Kapu
- Gyakoriság X Egyediségi Pont X Kapu

$p < 0,001$

Összeségében

- Egy szótag elég a felismeréshez pl. megfelelően a kohorsz elméletnek
- Minél korábbi az egyediség, annál korábbi a felismerés
- Gyakoriság segít
- Ritka szavak inkább alulról-felfelé
- Nyelvtani és fonotaktikai megszorítás is segít

Gating és entrópia

- **Prefixtypeoccurenceslog:** Az adott prefixummal kezdődő szótárban előforduló szavak számának 2-es alapú logaritmusa
- **prefixfreqlog:** Az adott prefixummal kezdődő, szótárban előforduló tokenek számának 2-es alapú logaritmusa
- **entrópia** – A webkorporusz feltételes entrópiája, adott prefixum feltételezése mellett. Informálisan, a fennmaradó bizonytalanságunk mértéke, amikor egy ismeretlen szó egy adott kezdőszeletét már megismertük.

$$H(W|x) = \sum_{w \in W} p(w|x) \log p(w|x)$$

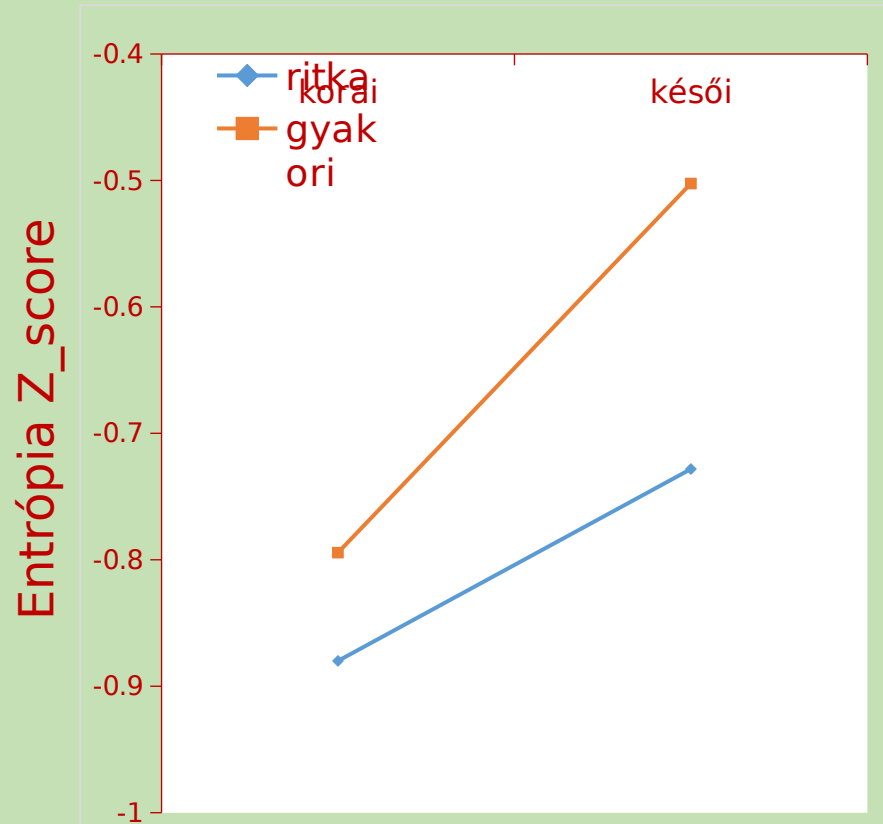
- **entropychange** – az entrópia csökkenése az előző kapuhoz képest (1. kapunál nincs értelmezve).
-

Felhasznált MOKK korpusz

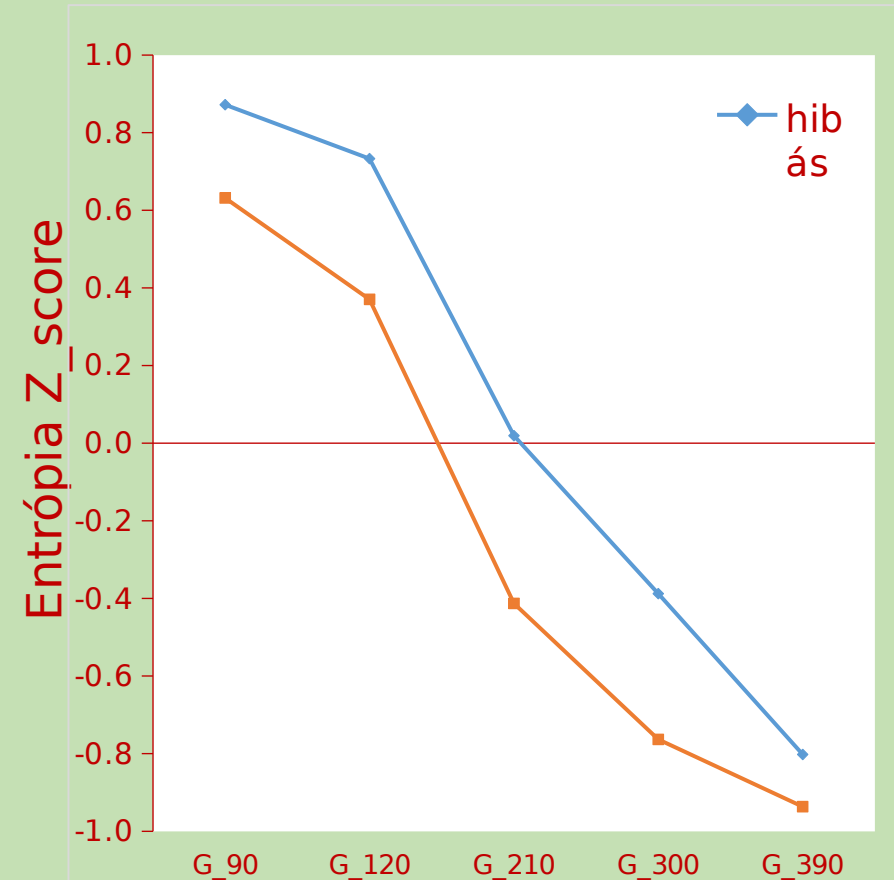


Korpusz	Oldal (millió)	token (millió)	type (millió)
Teljes	3,5	1486	19,1
60% Idegen kiiktat	3,125	1310	15,4
92% Csak diakritikus	1,918	928	10,9
96% Elütés mint a normálban	1,221	589	7,2

Gating és entrópia



- Egyediségi pont főhatás
- Gyakoriság főhatás
- Gyakoriság X Egyediségi pont interakció



- Gate főhatás
- ACC főhatás
- Gate X ACC interakció

$p < 0,001$

II. Kísérlet

- **GATING**

- Megszorítás Főnév
- megszorítás nélkül (Instrukció)

- **„On-line”** feldolgozáshoz közelebb áll

- RT (response box – mikrofon)
- Konfidencia-ítélet (1-2-3-4)

- **Résztvevők**

- 14 fő (7 férfi (28,16 év, $SD=14,05$), 7 nő (28,0 év, $SD=13,14$))

- **Ingerek**

- Ugyanaz, mint az 1. kísérletben



Az instrukció

Szavak részleteit fogod hallani. A feladatod, hogy kitaláld, mi lehet a szó.

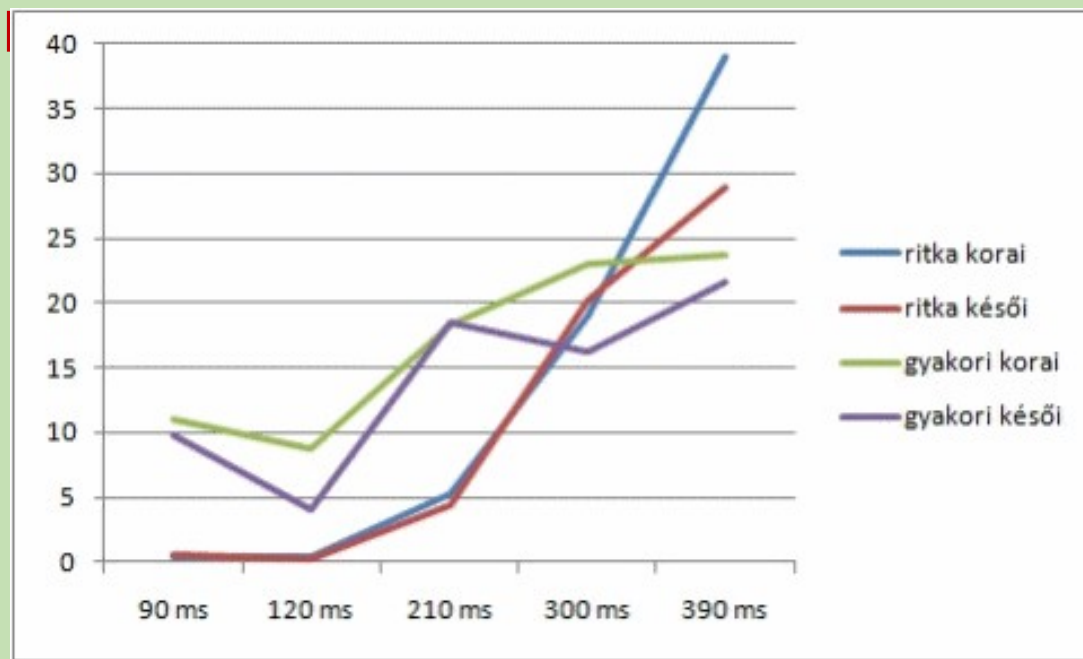
Azt is meg kell mondd a válasz után, hogy mennyire vagy biztos abban, hogy a tipped helyes volt. Minden elhangzott részlet után válaszolnod kell és biztossági ítéletet kell hoznod, de mindig csak az aktuális elhangzott részletre vonatkozóan.

Indulhat a feladat?

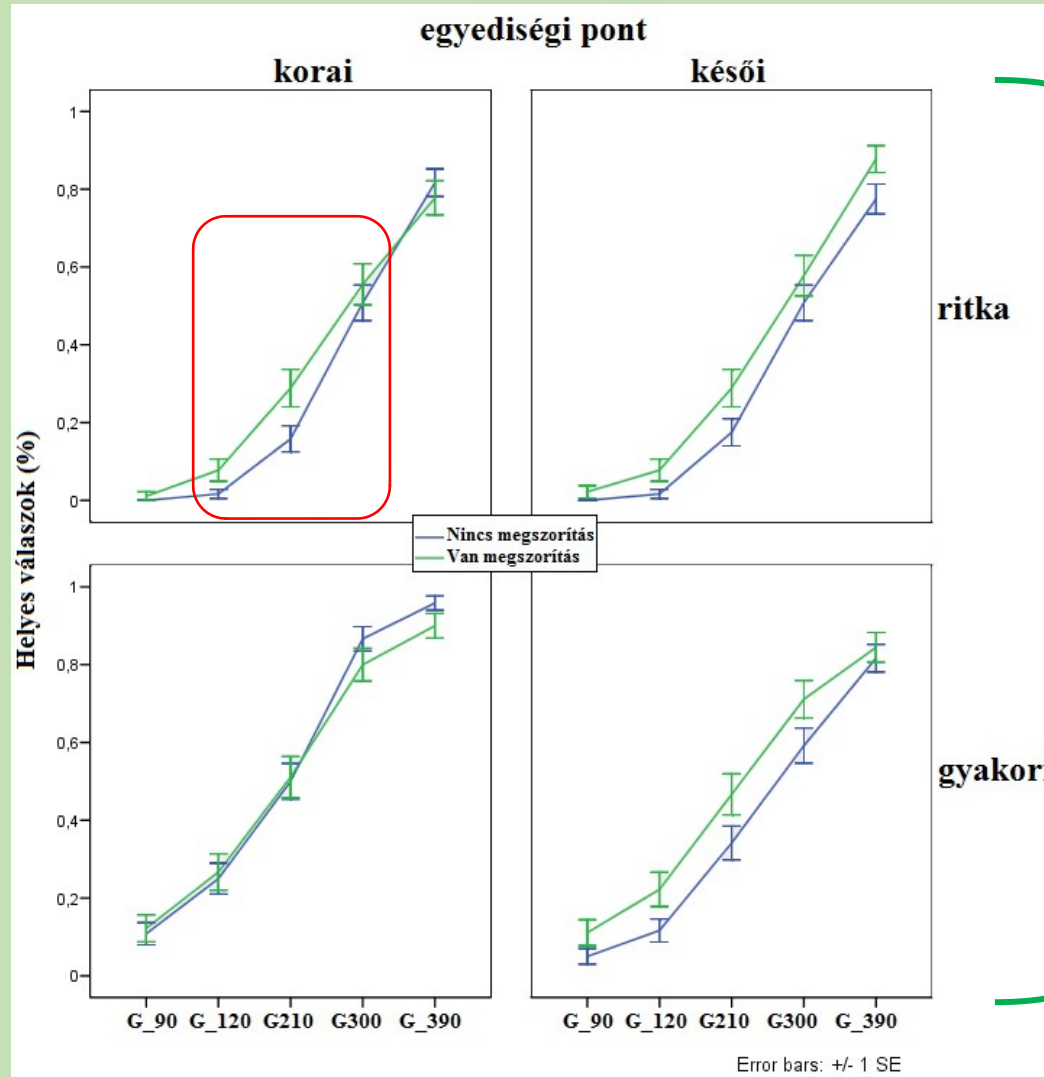
(Tedd fel a fülhallgatót!)

Felülről-Lefelé hatások

- Le lehet-e szűkíteni egy adott csoport a keresést?



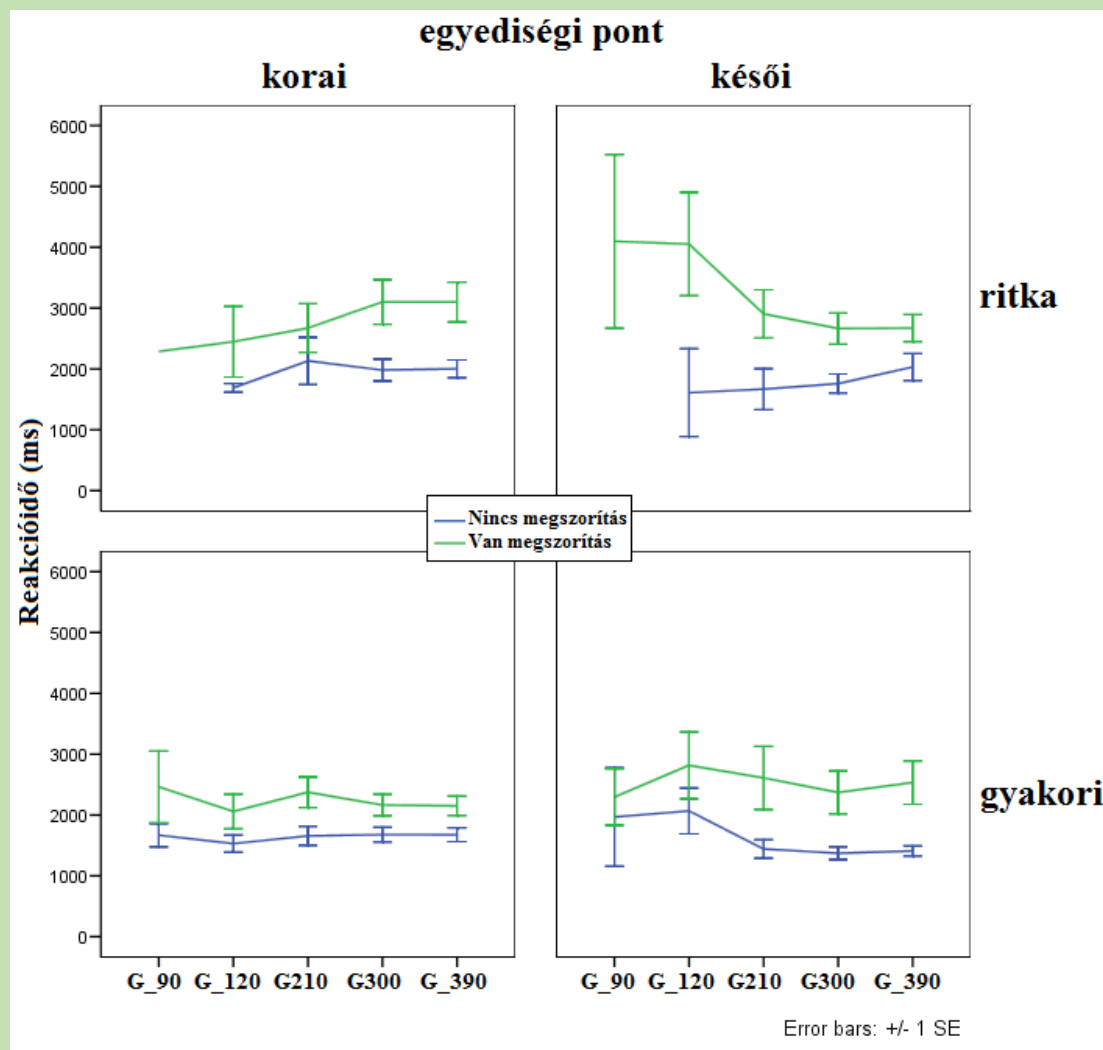
Pontosság



- A korai egyediségi pontúak esetén a kisebb kapuknál segít, a megszorítás

- A megszorítás a pontosságot a késői egyediségi pontú ingerek esetén növeli (gyakoriságtól függetlenül)

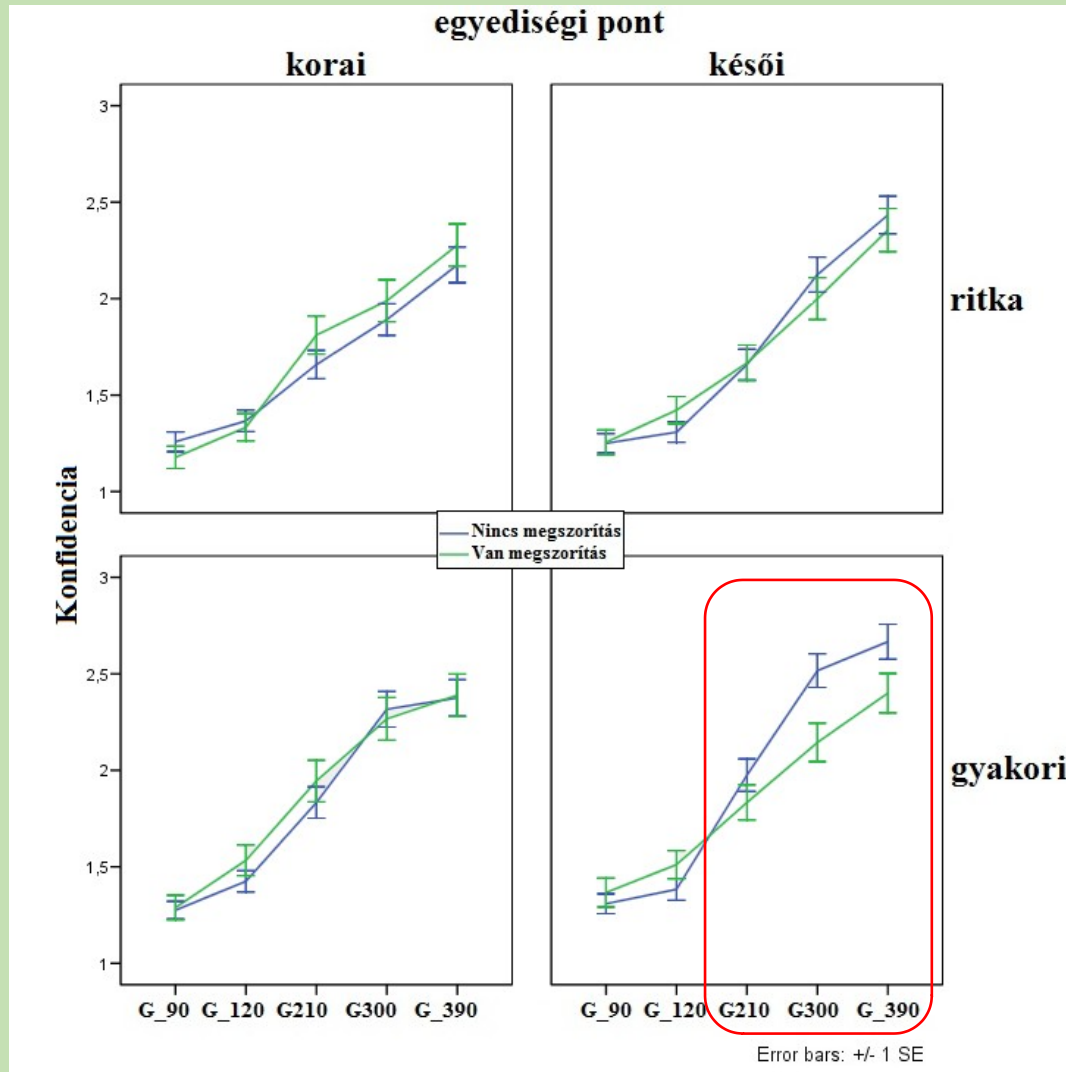
Reakcióidő



- a megszorítástól lassabbak lesznek a reakcióidők függetlenül a gyakoriságtól és az egyesíési ponttól.

- a ritka szavaknál egyesíési ponttól függetlenül a legkisebb kapu esetén nem volt helyes válasz – így itt nincs RI sem.

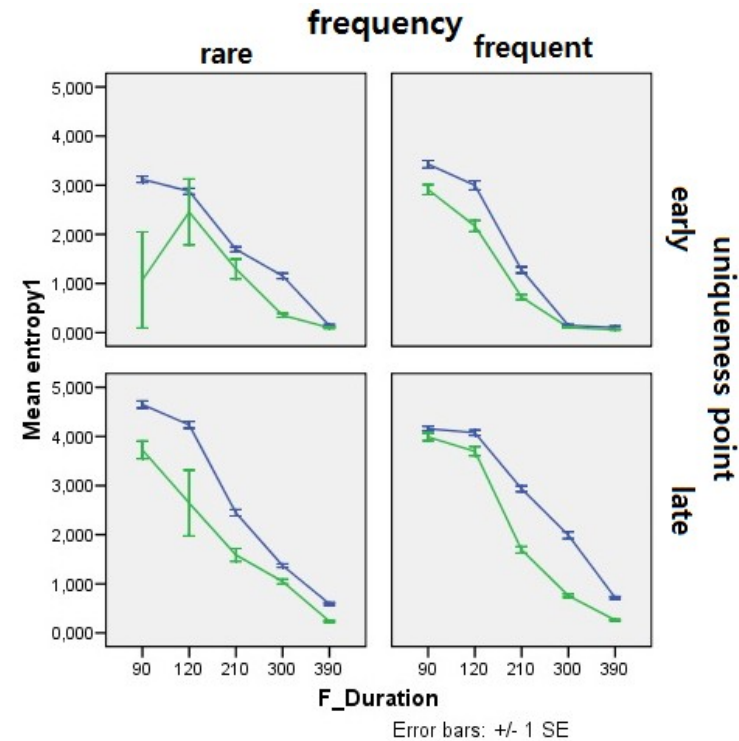
Konfidencia



- A megszorítás a gyakori-késői egyediségi pontú szavak esetén a hosszabb kapuktól (210+) kezdve csökkenti a biztossági ítéletet

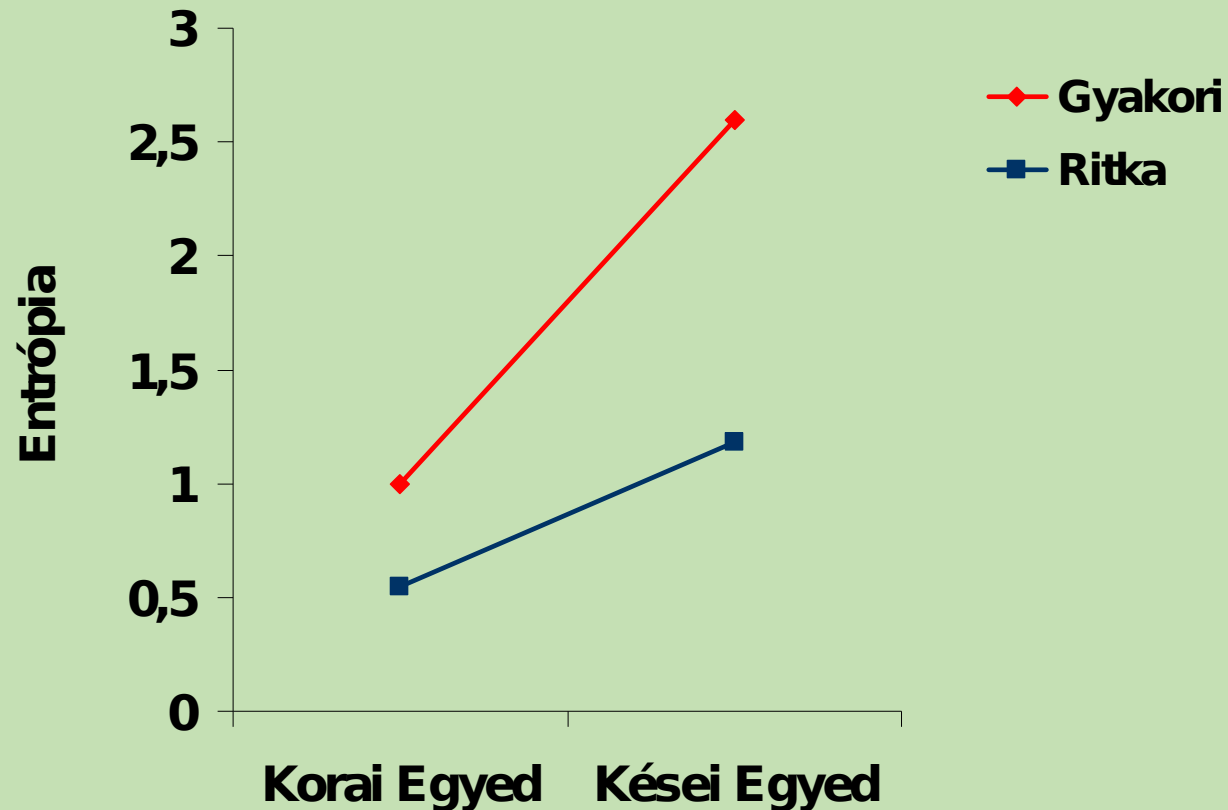
Entrópia és egyediség i pont

Entrópia akkor is
szignifikáns, ha az
egyediségi pont és a
gyakoriság ellenőrzött



Entrópia magasabb a gyakori szavaknál

Egyediségi pont: csökkent entrópia



Gating

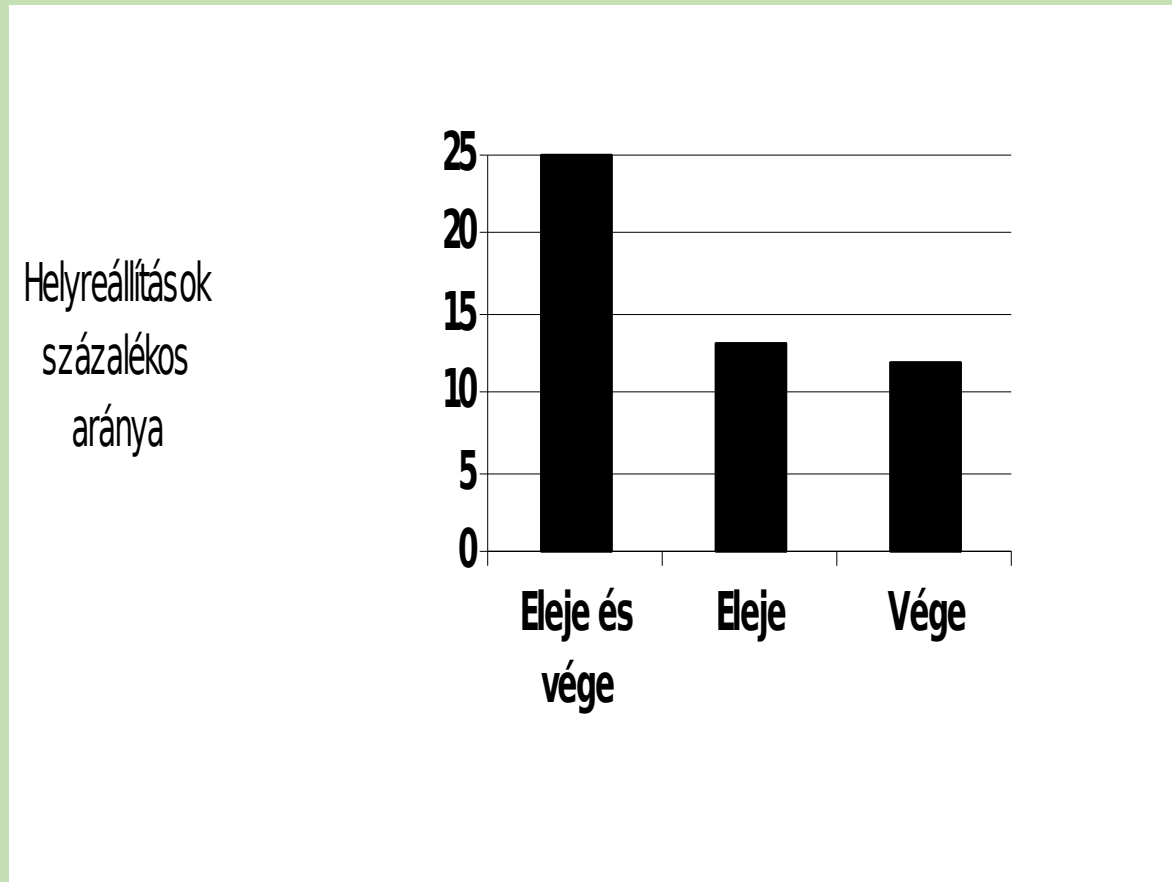
- Az entrópia jó közelítője a szófelismerés pontosságának
- A top-down megszorítás növeli a pontosságot, de lassítja a válaszadást
- Reakcióidők összevetése az entrópia-mutatókkal

A szavak eleje számít

Fazekas szövegjavítási feladat

Eleje - vége jó	Eleje jó	Vége jó
Ha egy pinaglló mbengbereti a sznyráát Pgenbiken, az akár tdoánrot is getheszejrt Amikerában. Vinalgkábun mdenin mninnedel öfüsszegg, menidn cseketüedlenk abápjalan vtja áltoztatha meg a jvöőt. Még elzépeklni is reómos	A nő nyumlogaar vákgyi. A házsagaás elrtotmol, a láany kamkdioasz- ő eregy kevbéés tajallá a hetlyé, és úgy ézir, rdviöidöre mujszá tálvo keneirlü a hétkölótpaijzan . Eztré fodjaag el egy bajatrá megtáváhí, aknike van egy	A ágmany acsk sesűrkeéget üszl és aszőrekot. Max enm thite nolva, ohgy a ajsát őrbén lkel taszegpamtaln ia zet. Jó arzsu olvt, kai glodoban él a leséefge és álksinya lelmett. A

A szavak rekonstrukciója sikeresebb, ha mind a szó eleje mind a vége helyesen van leírva



Az előkészítési helyzet

RPROBLÉMA

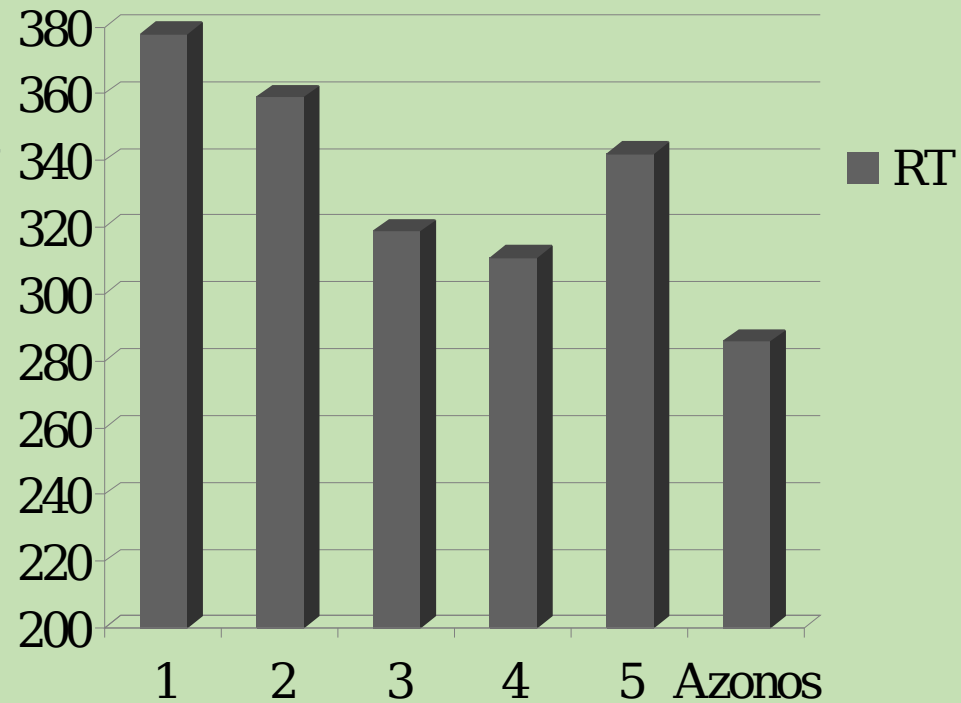
PROBLÉM

A



Előfeszítési hatás a kezdetnél: A rontott szó lelassít az elején és a végén

Várhelyi Klára



PROBLÉMA
PORBLÉMA
PROLBÉMA
PROBLMÉA
PROBLÉAM

Összefoglalva

- A szókezdet kiemelkedő jelentősége a hozzáférésben a magyarban is nyilvánvaló
- A szófelismerés érzékenyebb az entrópia értékekre és az alaktani szerkezetre mint magára a gyakoriságra
- Az entrópia változás fontos a szomszédsági hatások értelmezésében

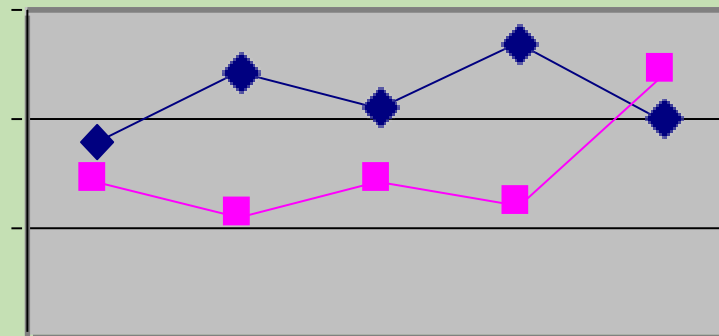
Az alaktan kérdései

- Szegmentálást segíti: hangsúly elől, szóalak korlátok
- A töveket és a ragot dolgozzuk fel elsődlegesen (Juhász és Pléh), lassabban, mint az angol vagy olasz
- A végződés feldolgozás már menet közben, hallgatáskor végbemegy (Gergely)

Elemző
ő
default

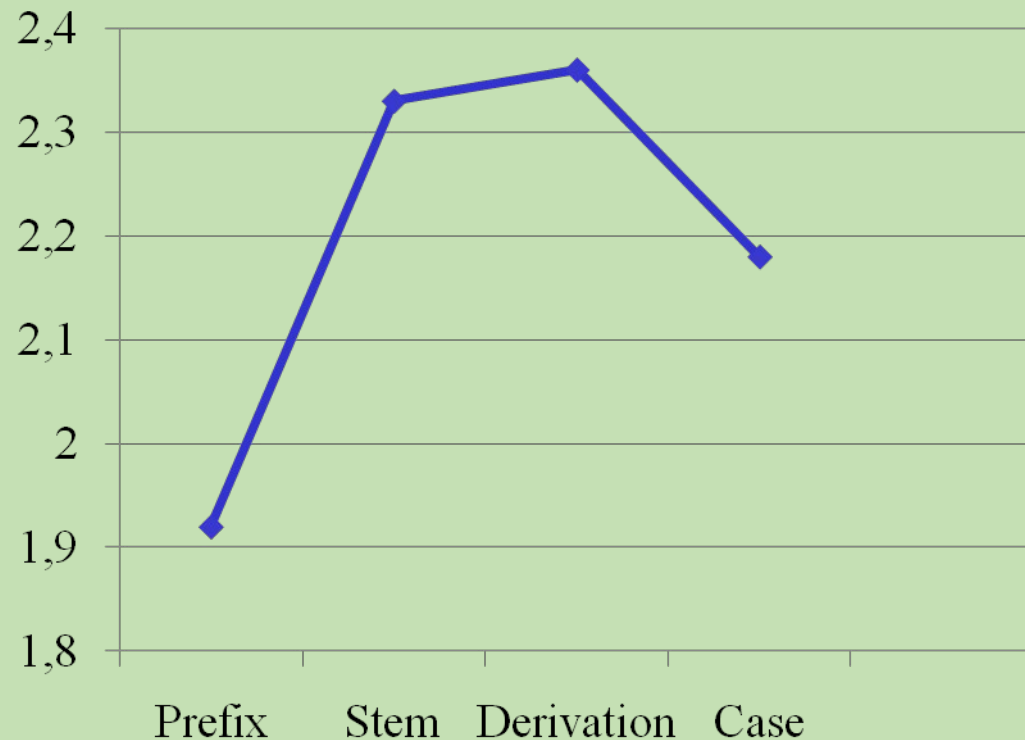
Egészleges Gyakori
Egyedi:
házasság

**Irreleváns kattánások
felismerése mondat
hallgatása közben**



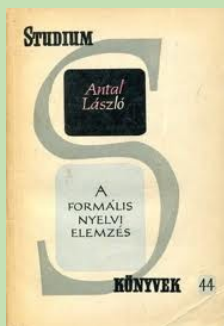
csók-ja-id-tól
Levedula

Döntések hosszú sok morfémás szavaknál: Pléh és Juhász 1995

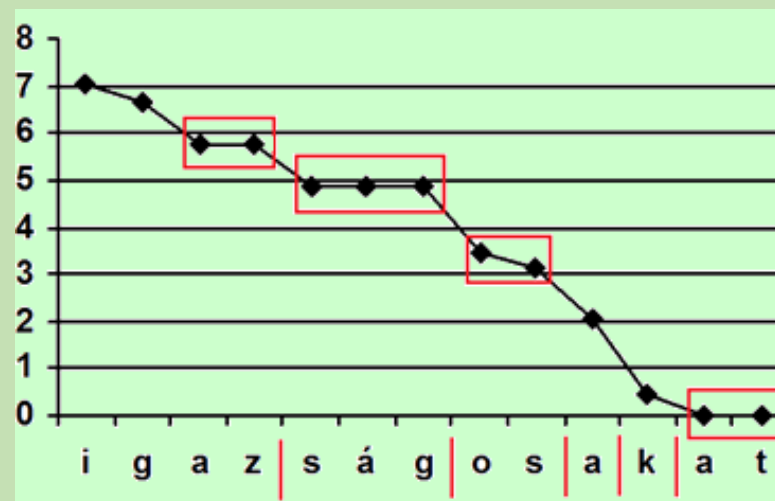


bathtub effect
Aithchison

Korai felvetések az információs alapú morfológiáról

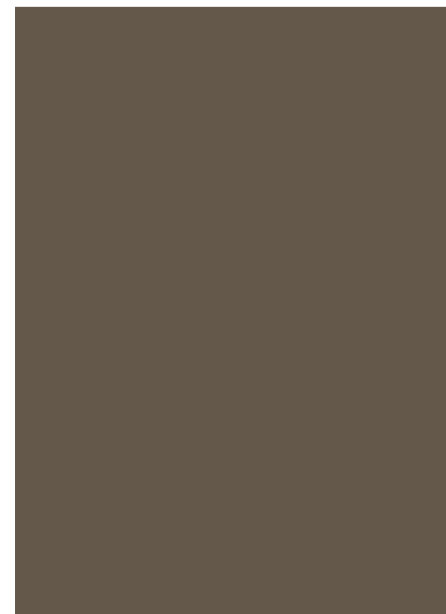


- Antal László (1964) a szó testében az általános tendencia az entrópia fokozatos csökkenése. Morfémahatáron megszakad ez a csökkenés
- *igaz-ság-os-ak-at*



Szisztematikusa
bban Hasonló
szavak, mint a
gatingnél

- Böllár
Tő
- Böllér-**a**k böllér**u**k
Jel
- Böllér-**n**a**k** böllér-
nu**k** Rag
- Böllér-**a**k-**n**e**k** böllér-
uk-**n**e**k** Jel belül



A vizsgálati anyag alapja

- 60 kétszótagú főnév, Németh Kornél gating vizsgálata alapján
- Forrás: MOKK korpusz (www.szoszablya.hu)
- 4 csoport egyediségi pont és gyakoriság alapján

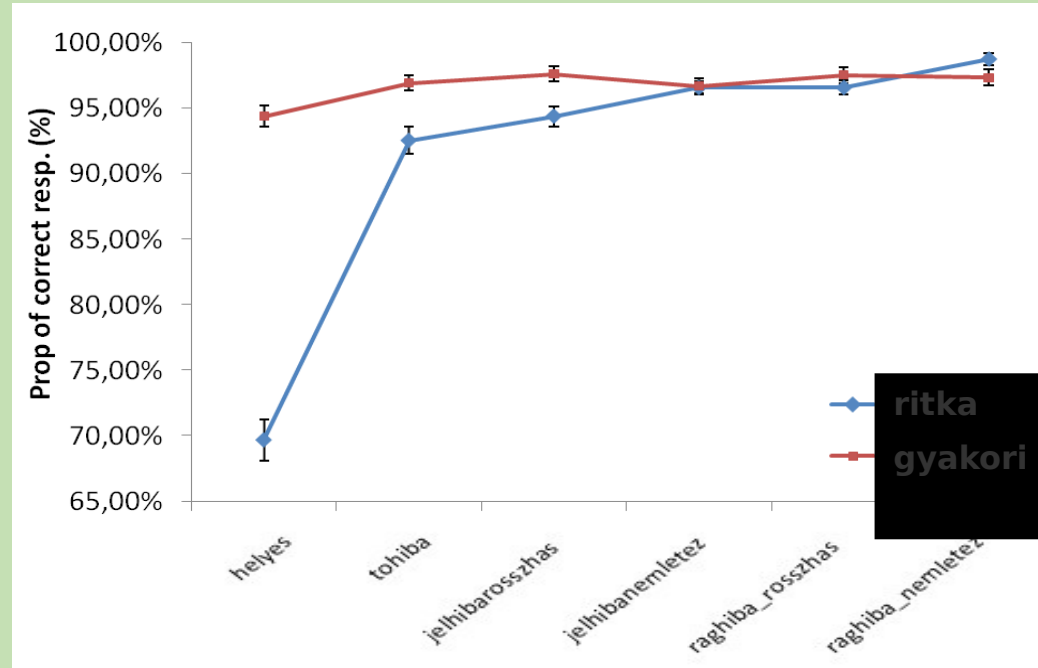
	Gyakor i	Ritka
Korai	kenyér, műsor	böllér , gyűszű
Késő i	város , csapat	pincsi , kroket

A ragozott szavak

- Előfordulások: csak tő, jel, rag vagy jel és rag
- Rontás a tőben (kunyérem) a jelben (kenyérid) vagy a ragban (kenyeredbun)
- Létező, de rosszul illesztett (kenyérim) vagy nem létező toldalékok (kenyérum)

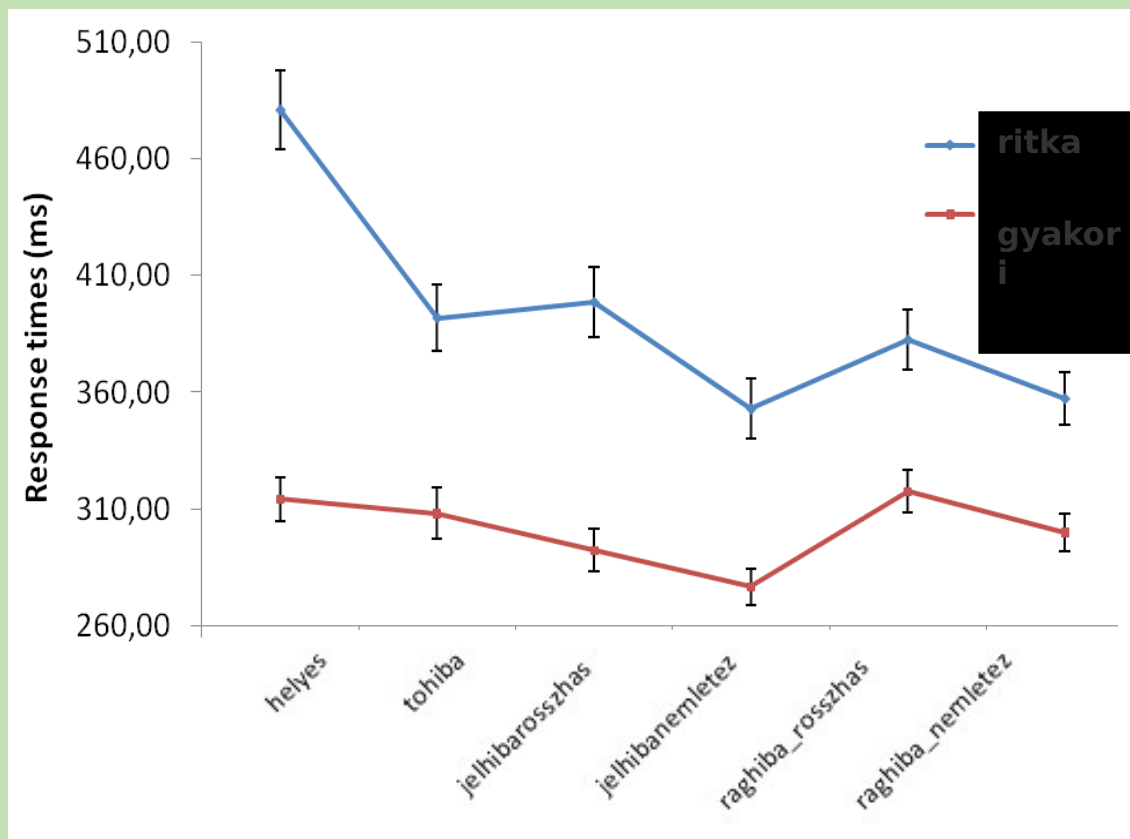
kenyérnek		kenyérben		kenyérr e	
kenyerek	kenyerak/ kenyerik	kenyerem	kenyeram/ kenyerim	kenyer ed	kenyerad/ kenyerid
kenyerekn ek	kenyerakne k/ kenyerikne k	kenyerekb en	kenyerakb en/ kenyeriken	kenyer ekre	kenyerakr a/ kenyerikra
kenyeremn ek	kenyeramn ek/ kenyerimne k	kenyeremb en	kenyeramb en/ kenyerimb en	kenyer emre	kenyeram ra/ kenyerimr e
	kenyeradn		kenyeradb		kenyeradr

Eredmények – helyes válaszok



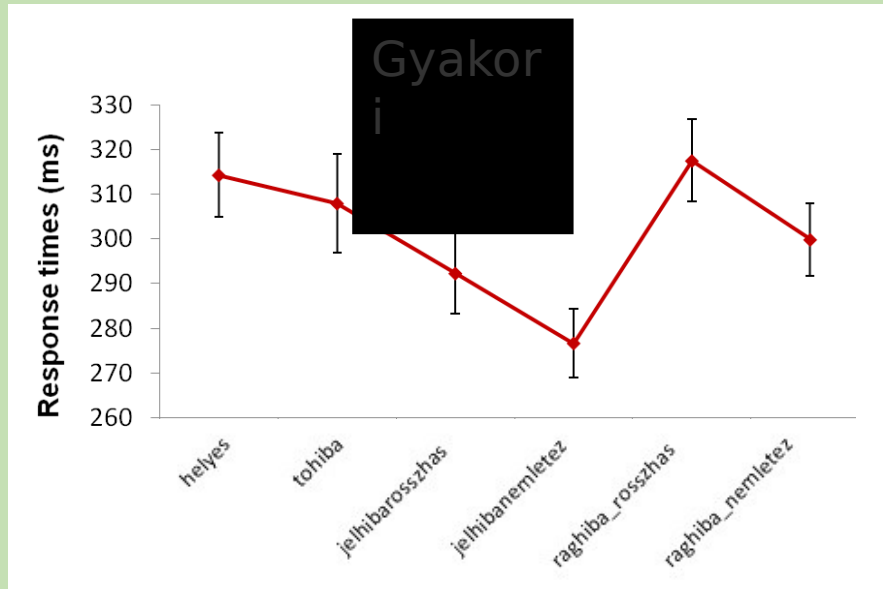
- Mind a gyakoriság mind a hibatípus hatása szignifikáns
- A szó gyakorisága és a hibatípus felismerésének gyakorisága között is szignifikáns korreláció figyelhető meg
- Minél később fordul elő a hiba a szóban, elutasítása annál könnyebb és gyorsabb

Eredmények - reakcióidők

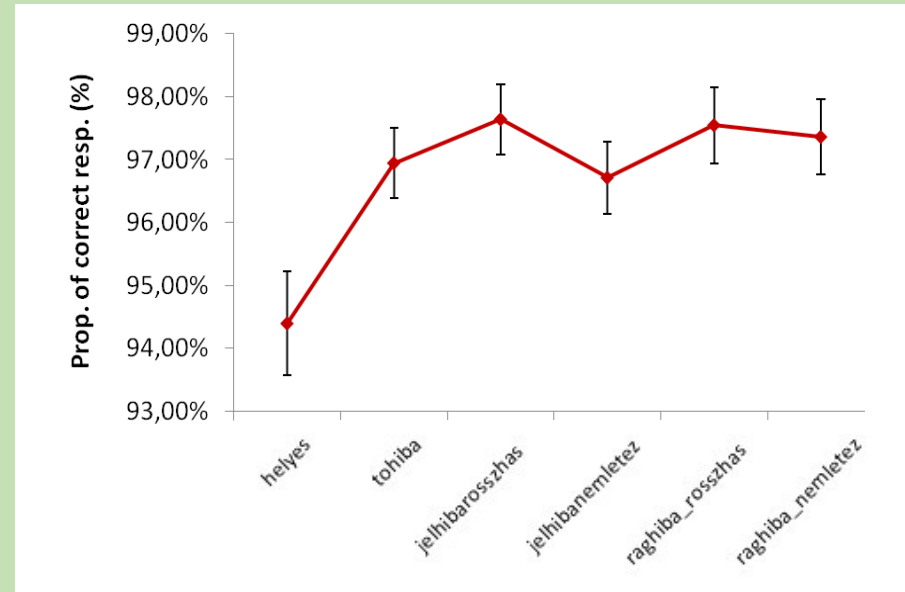


- Mind a gyakoriság, mind a hibatípus hatása szignifikáns
- A létező szavak elfogadása lassabb, mint a nem létezők elutasítása
- A nem létező tövek elutasítása lassabb, mint a nem létező toldalékoké
- Nincs egyértelmű fürdőkád hatás
- A két magánhangzó harmóniát sértő hibatípus közt nincs különbség, viszont ezekhez képest a nem létező toldalékok elutasítása gyorsabb

Szóközépén előforduló rontások gyakori szavaknál



Reakcióidő



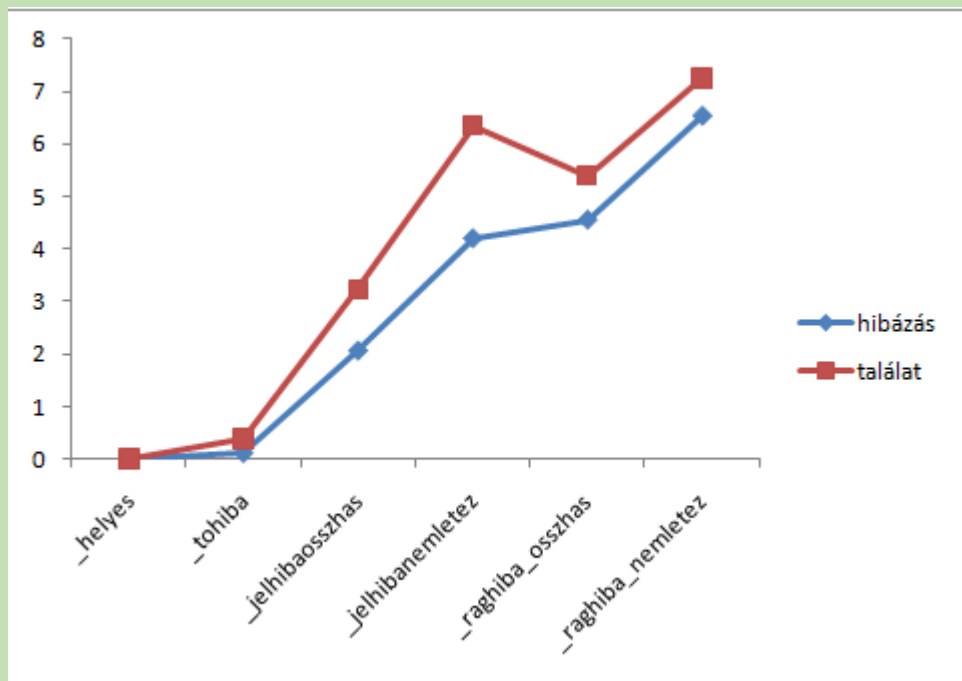
Helyes
válaszok

- Gyorsabb reakcióidő, de gyakoribb hibázások
- Lehet egy optimalizálási mechanizmus eredménye

A rontás alattomoss ága

- A rontás előtt lévő 4 karakter előfordulási gyakorisága a MOKK korpuszban, pl.: bölléred - böllérud
léru-782 előfordulás
lére-75283 előfordulás
- A két szám hányadosának logaritmusa (ngram-faktor) határozza meg a rontás alattomosságának a mértékét

Eredmények – a rontás alattomossága



A rontás alattomossága minden hibatípus esetében segít a rontások felismerésének bejósolásában

- Az ngram-faktor fordítottan korrelál az alattomossággal
- Minél ritkábban fordul elő a rontás előtti betűnégyes az eredeti betűnégyeshez képest, annál kevésbé alattomos a rontás
- **Minél kevésbé alattomos a rontás annál nagyobb eséllyel veszik észre**
- A kategoriánként elvégzett repeated measures Anova formálisan is alátámasztja a hipotézist

Alaktani és lexikai entrópia

- Kabát –

- ok

- -ot

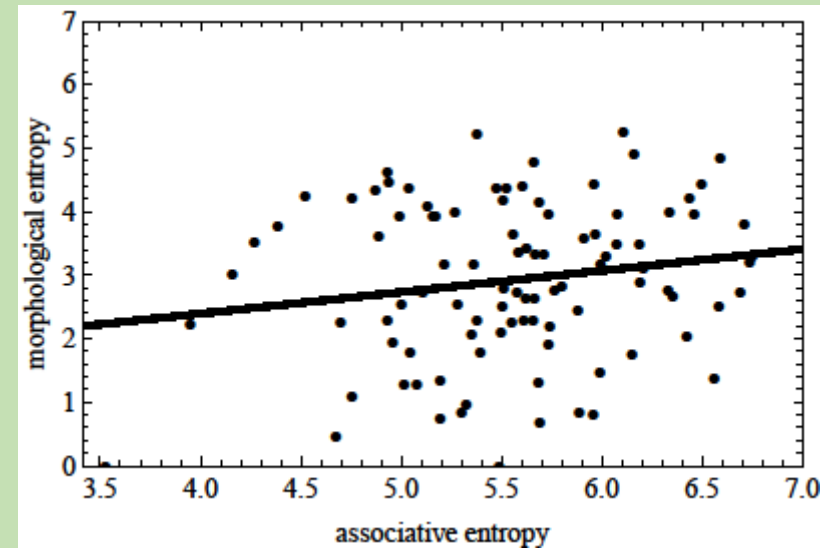
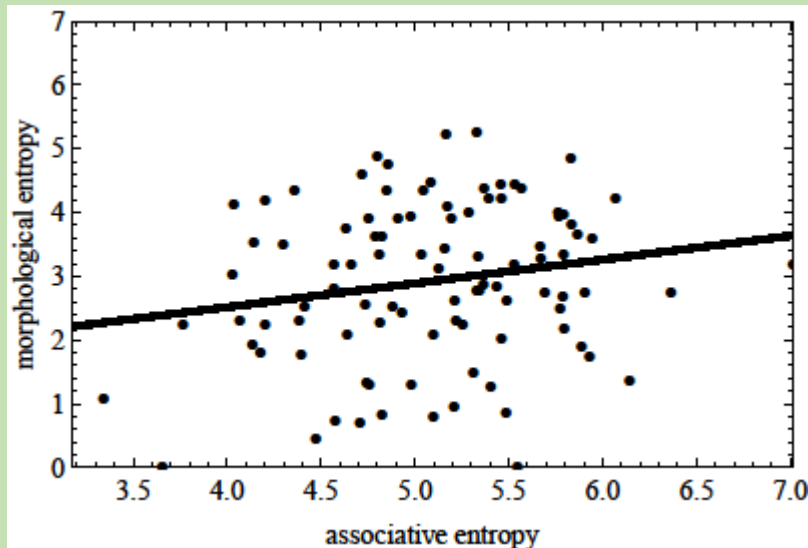
- -jaimat

- Kabát kalap

- fogas

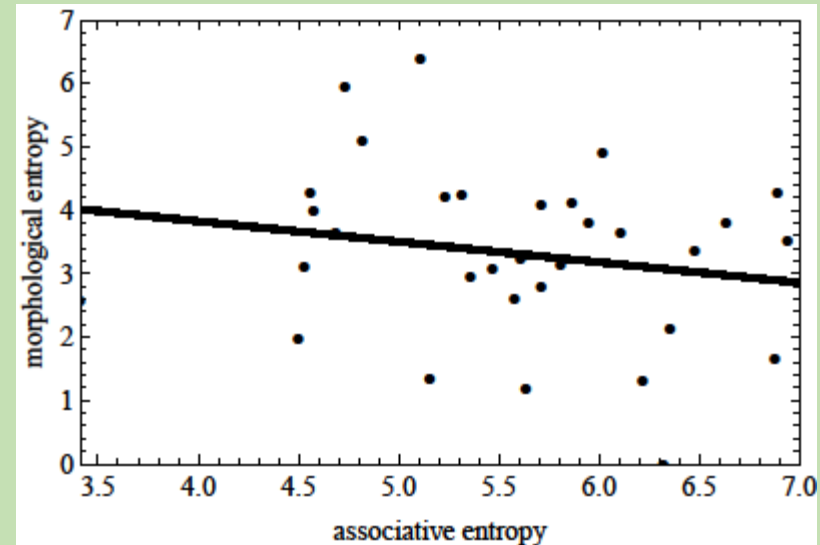
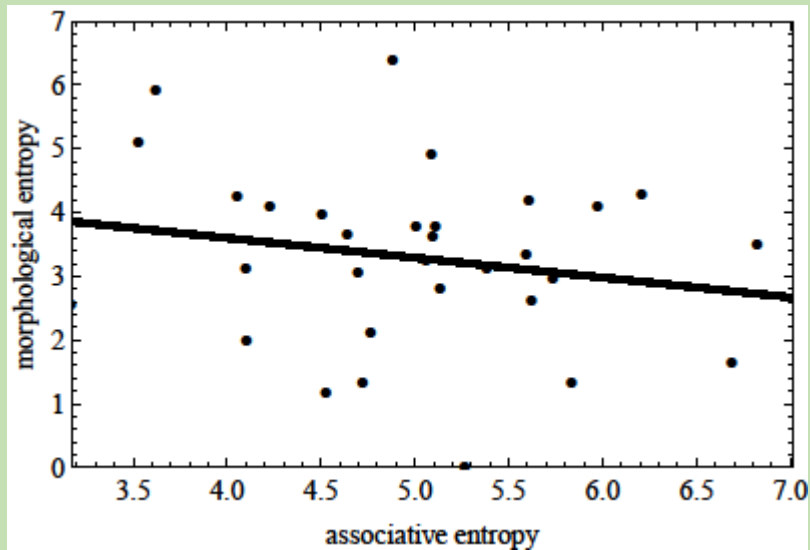
- szék

Gyenge pozitív korreláció főneveknél 10-14 20-24 A kettő viszonya Czégel, Lengyel és Pléh



Gyenge negatív korreláció igéknél

10-14 18-24

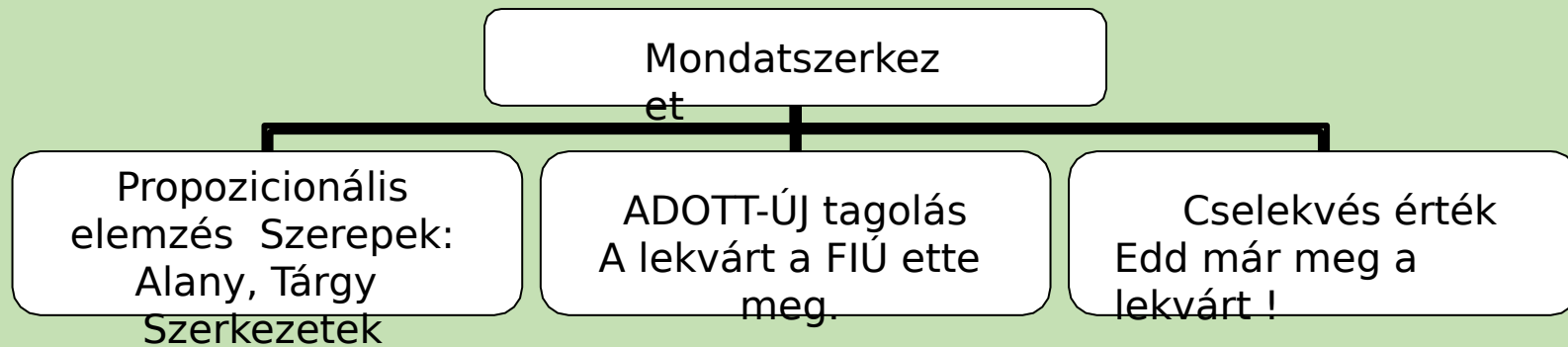


	10-14	18-24
Nouns	0.20	0.18
Verbs	-0.19	-0.19

Összefoglalás

- A gyakoriság és a hibatípus hatása és ezek kapcsolata mind a helyes válaszok, mind a reakcióidők tekintetében szignifikáns
- Minél később fordul elő a hiba a szóban, elutasítása annál könnyebb és gyorsabb
- A nem létező toldalékok elutasítása lassabb
- Minél alattomosabb egy hiba, annál kisebb eséllyel kerül elutasításra

A mondat megértés kérdései



Egy mondat
RSVPvel

A
magas

királyn
ő

kizsákmányol
ta

a
vacogó

rendőrök
et

a
kertben.

Ellenőrzé

S

- Királyn
ő

- ker
t

A propozicionális elemzés kérdései

- Mondatbeli szerep eldöntése:
mit használunk? Szórend,
rag, élőség
- Szabályok és Statisztika
viszonya
- Tévedések a gyors
elemzésben
- Nyelvek eltérései
- A mondat szerkezet
felállításának elvei

Nyelvek
eltérései a
tényezőkben
Bates és
MacWhinney,
1998

- Egyszerű mondat
értelmezés

**The Competition Model (Bates
and MacWhinney, 1982)**

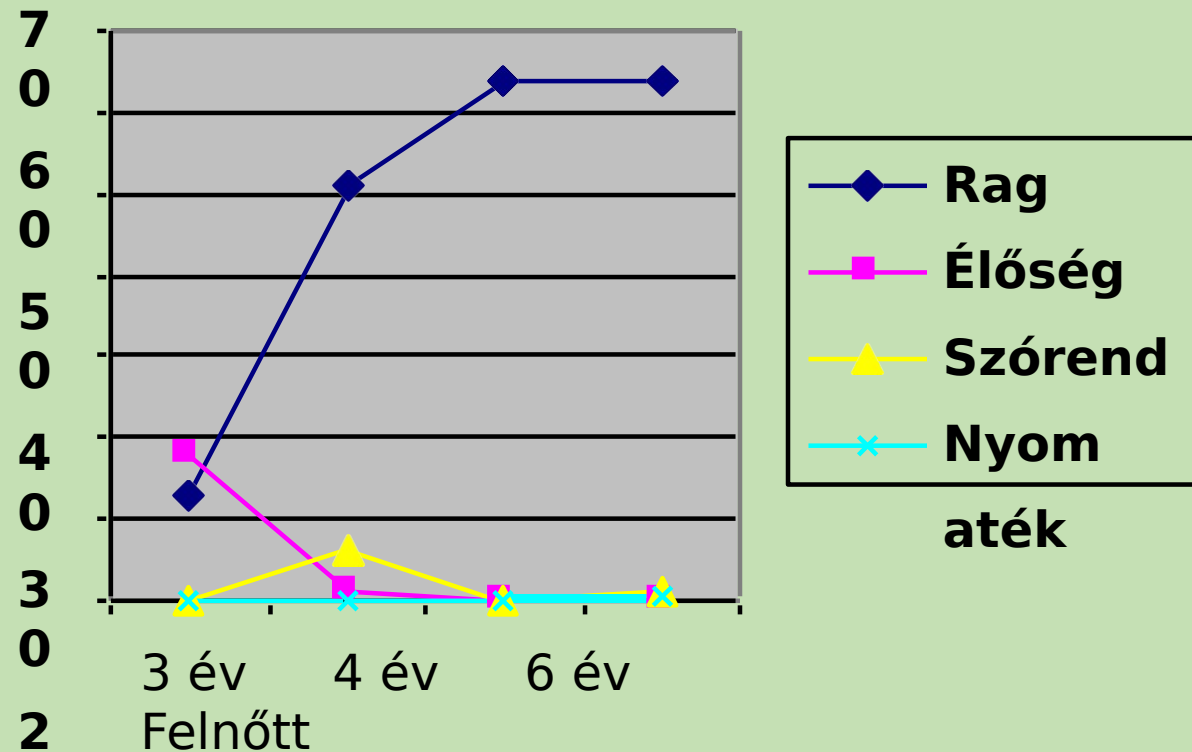
Language	Child	Adult
English	Order > Anim > Agr	Order > Anim > Agr
French	Order > Anim	Agr > Anim > Order
Dutch	Order > Case > Anim	Case > Order > Anim
Croatian	Anim > Case > Order	Case > Agr > Anim
Turkish	Case	Case > Anim > Order
Hungarian	Anim > Case > Order	Case > Order > Agr
Walpiri	Anim > Case > Order	Case > Anim > Order

A különböző tényezők magyarázó értéke magyar gyermekeknél

Megmagyarázott
variancia kergeti a
lovat



ló illetve
ceruza



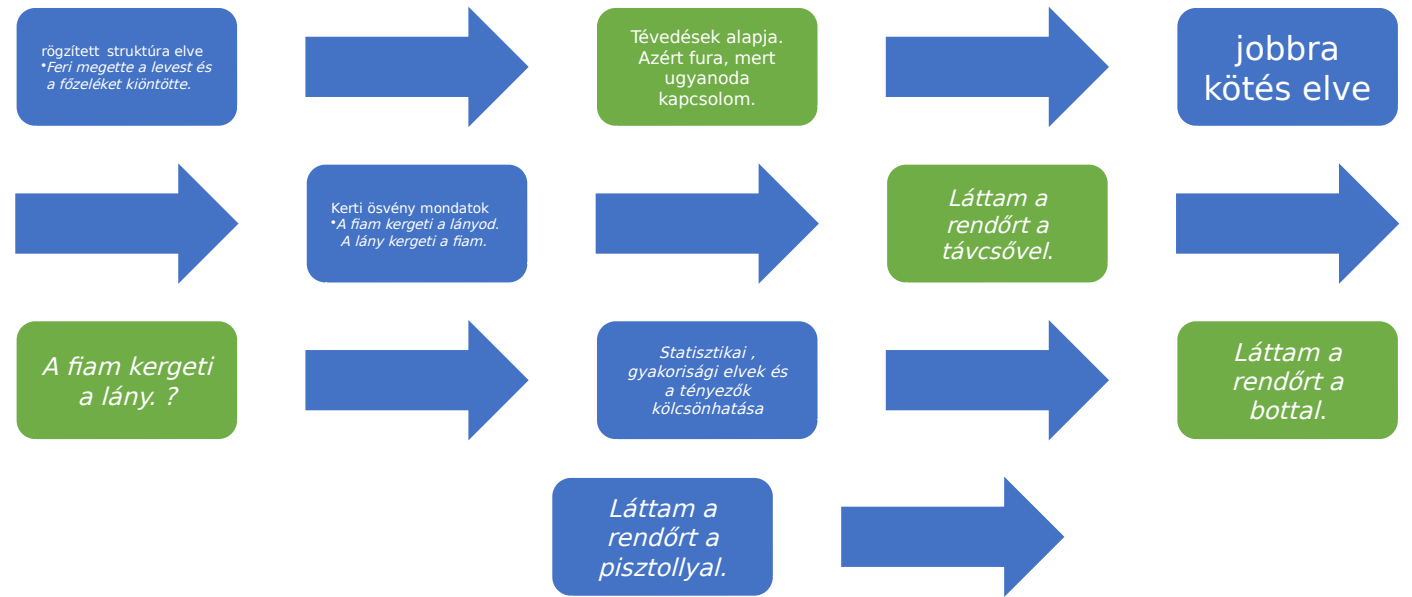
2
0
0

3 év 4 év 6 év

aték

A propozicionális

e
l
e
m
z
é
s



Elvárási hatás és entrópia a mondatmegértésben

Számos adat az ige és vonzat közti elvárásról

Lehetséges feltételes entrópia hatások

Játszadozás a társakKAL

Magyar ESZKÖZ hatások

Az ESZKÖZÖk entrópia függő értelmezése

Az vonzatelvárás feldolgozási hatásainak szakaszai

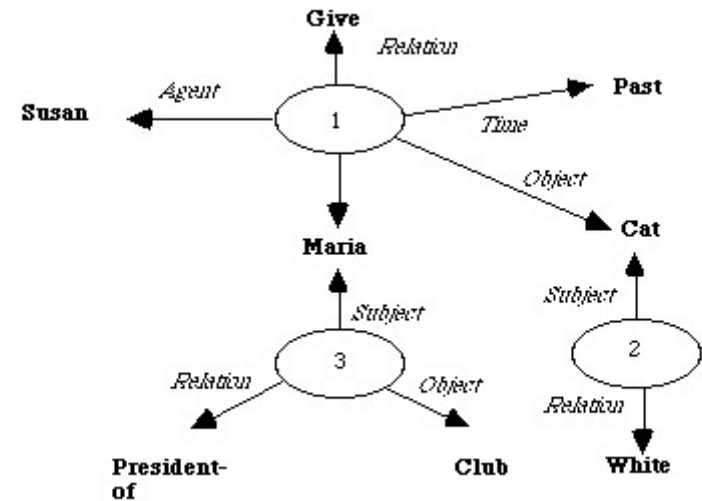
- Korai séma elméletek
- Kognitív elvárás

John gave the book to Mary.



- 1972 Schank

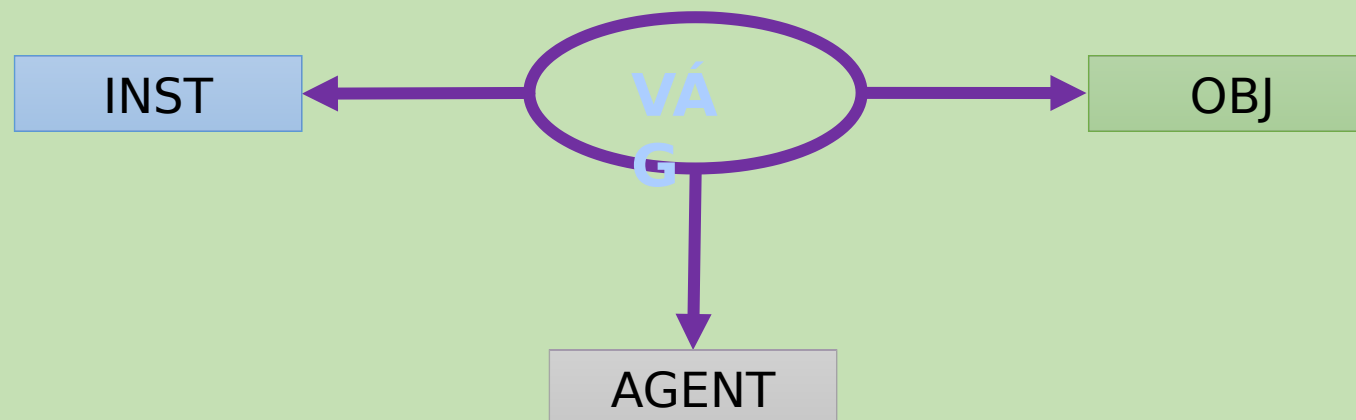
Kintsch 1974



A séma alapú feldolgozás alapgondolata

**Predikátumok
aktiválása az
emlékezetből**

**Vonzatok az üres
helyekhez rendelése**



Nyelvészeti gondolatmenet

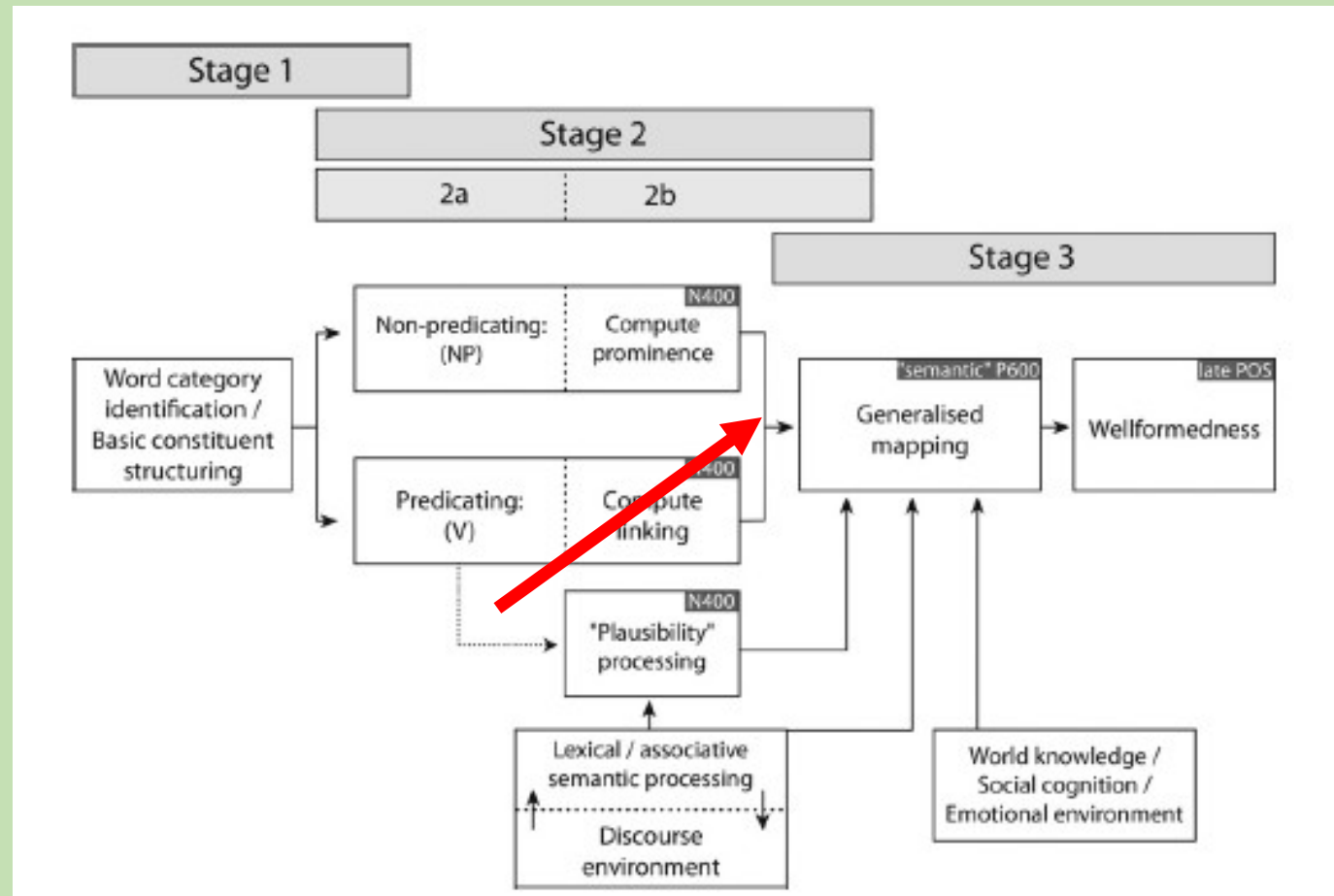
Tematikus szerepek

Kétértelműség a hozzárendelésben

*Láttam a rendőrt **a távcsővel.***

*Láttam a rendőrt **a pisztollyal.***

Bornkessel and Schmelevsky, 2006



Magyar érdekesességek

Vonzat
viszonyt a rag
jelzi

Alternatívák:
Gondolkozott a
hajón a lányon.

Többszörö
s keretek

Elvont viszonyok
helyi ragokkal.
Gondol a lányra

Vonzat
illetve hely.

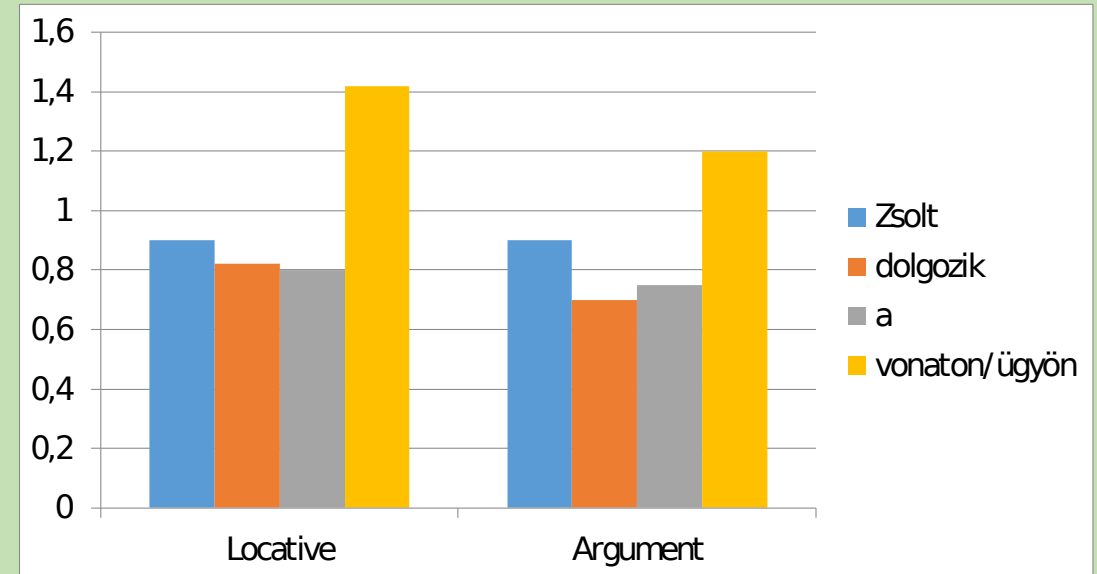
játszik tárgyatlan,
tárgyas, eszköz,
társ

Típus	Példa
Egyértelmű	Emlékszik a fiúra. Haragszik a tanítóra. Készül a versenyre. Találkozott a lánnyal. Csókolózott a rendőrrel.
Kétféle vonzat	Gondol a lányra. Gondol valamit.
Kétértelmű vonzat/határozó	Gondolkodik a lányon. Gondolkodik a hajón.

Változat
ok

Gervain és Pléh (2003) lokatív alteráció

- Főnév után kétértelmű vonzat/vagy hely lassabb olvasás



Fekete és Pléh 2011

Kétirányú: *János csókolódzott Marival.*

Egyirányú: *János kikezeett Marival.*

A kísérlet releváns igéi

komitative
(kétirányú)

borozott
találkozott

tegeződött
csókolózott
sétálgatott
bulizott
mulatozott
párbajozott
verekedett
énekelt

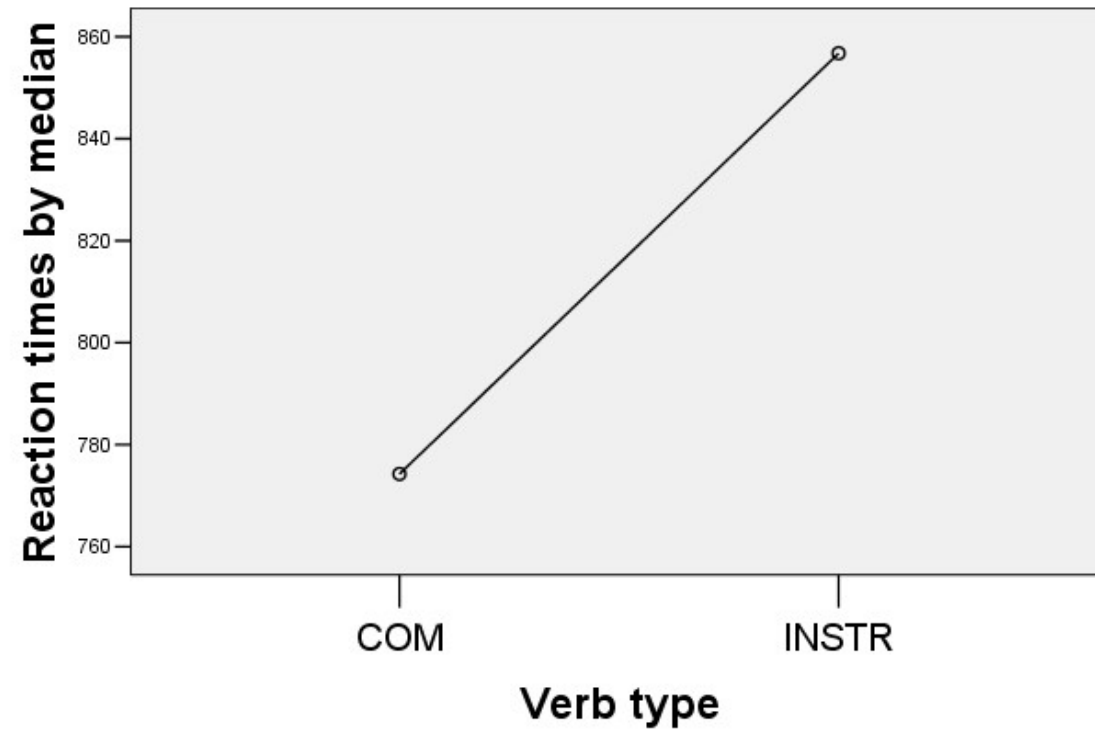
Eszköz (egyirányú)

csipkelődött
foglalkozott
Incselkedett
gúnyolódott
ordított
szimpatizált
csúfolódott
együttérzett
kikezdett
törődött

Teszt mondatok

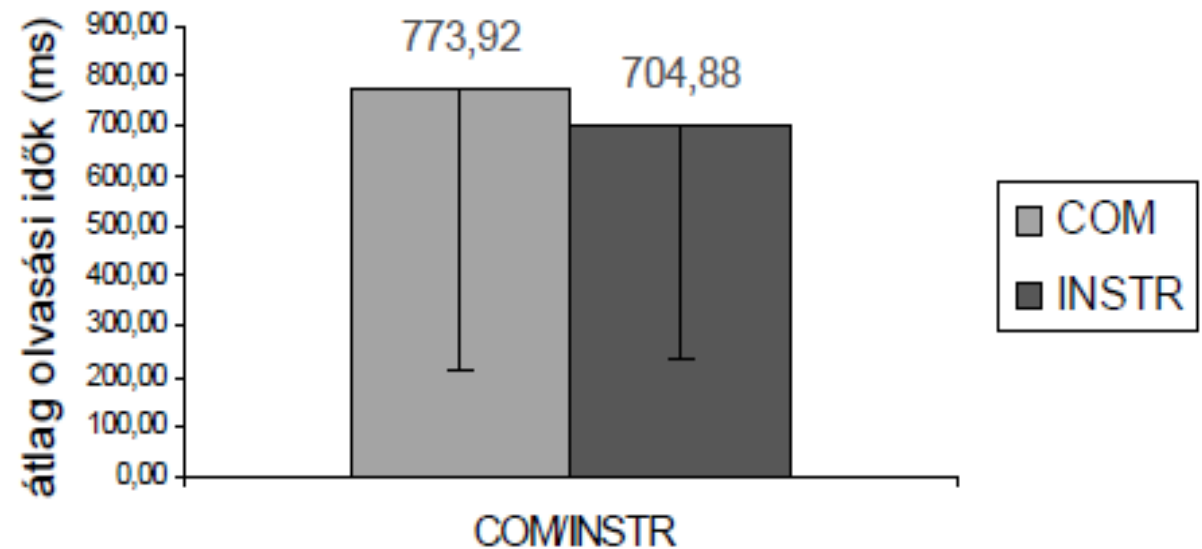
A lány a fiúval bulizott.	Összejöttek /plul/ másnap.	A lány a fiúval kikezdett.	Összejöttek /plul/ másnap.
	Összejött /sing/ vele.		Összejött /sing/ vele.

Main effect of verb type (word 5)



A vonzatot
egyirányú
ige után
gyorsabban
olvassuk

2. kísérlet (NVN): az ige utáni NP-k átlag olvasási időtartamai



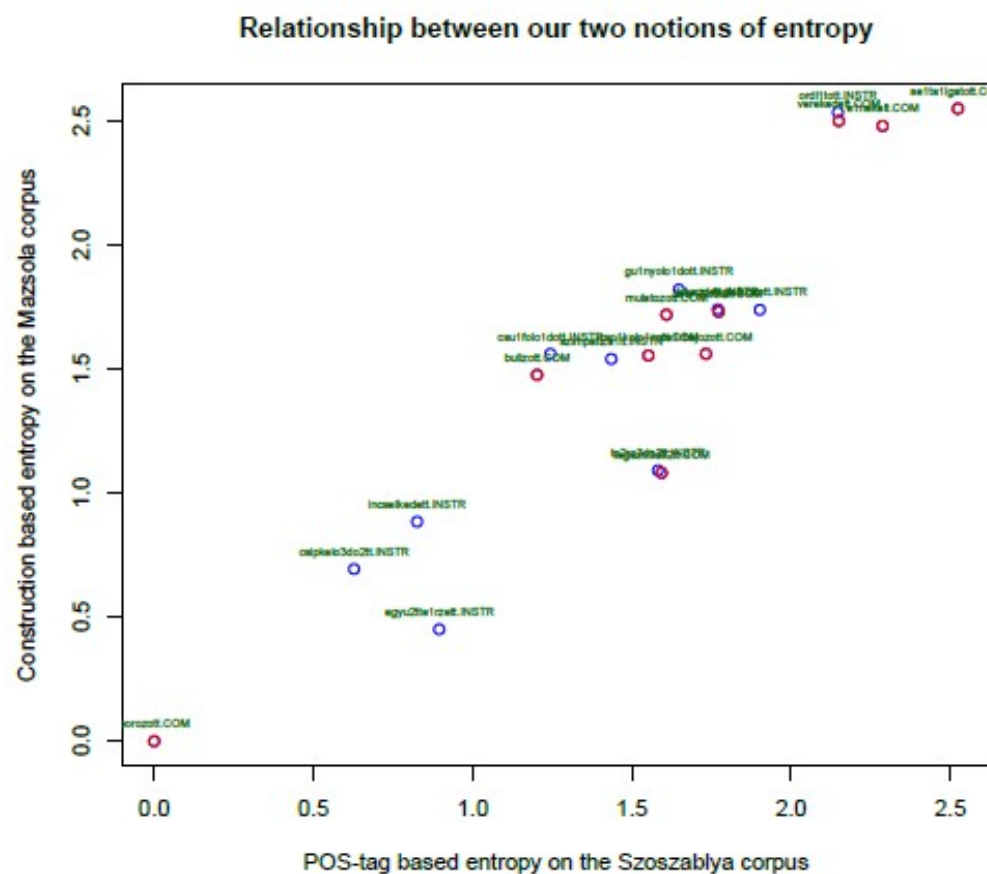
Az NP-k olvasási időtartamai a két
igetípus után

Újraelemzés az entrópia segítségével

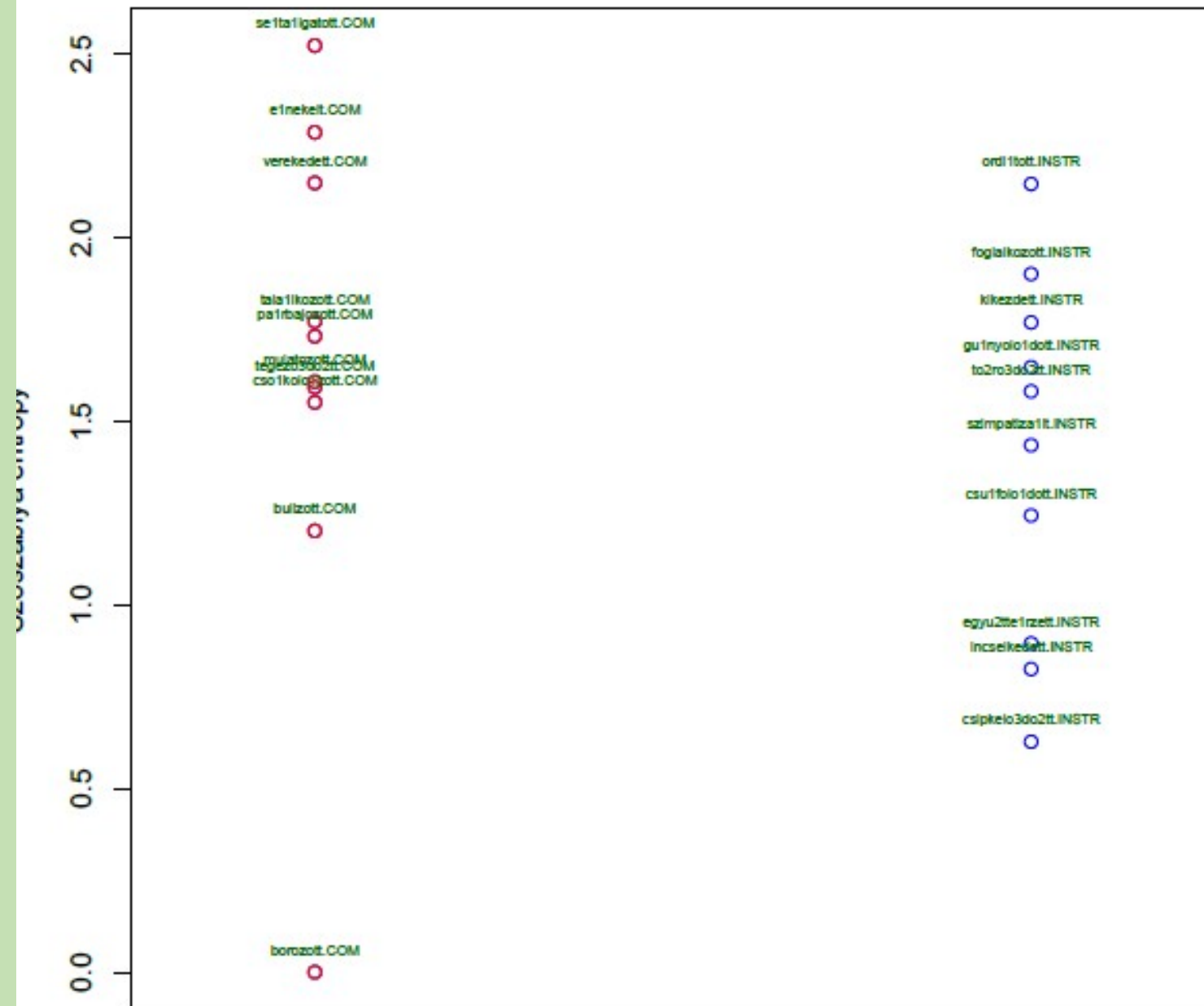
- Az ige entrópiája: mennyire követheti változatos esetkeretű főnév ?

$$H(W) = - \sum_{w \in W} p(w) \log_2 p(w)$$

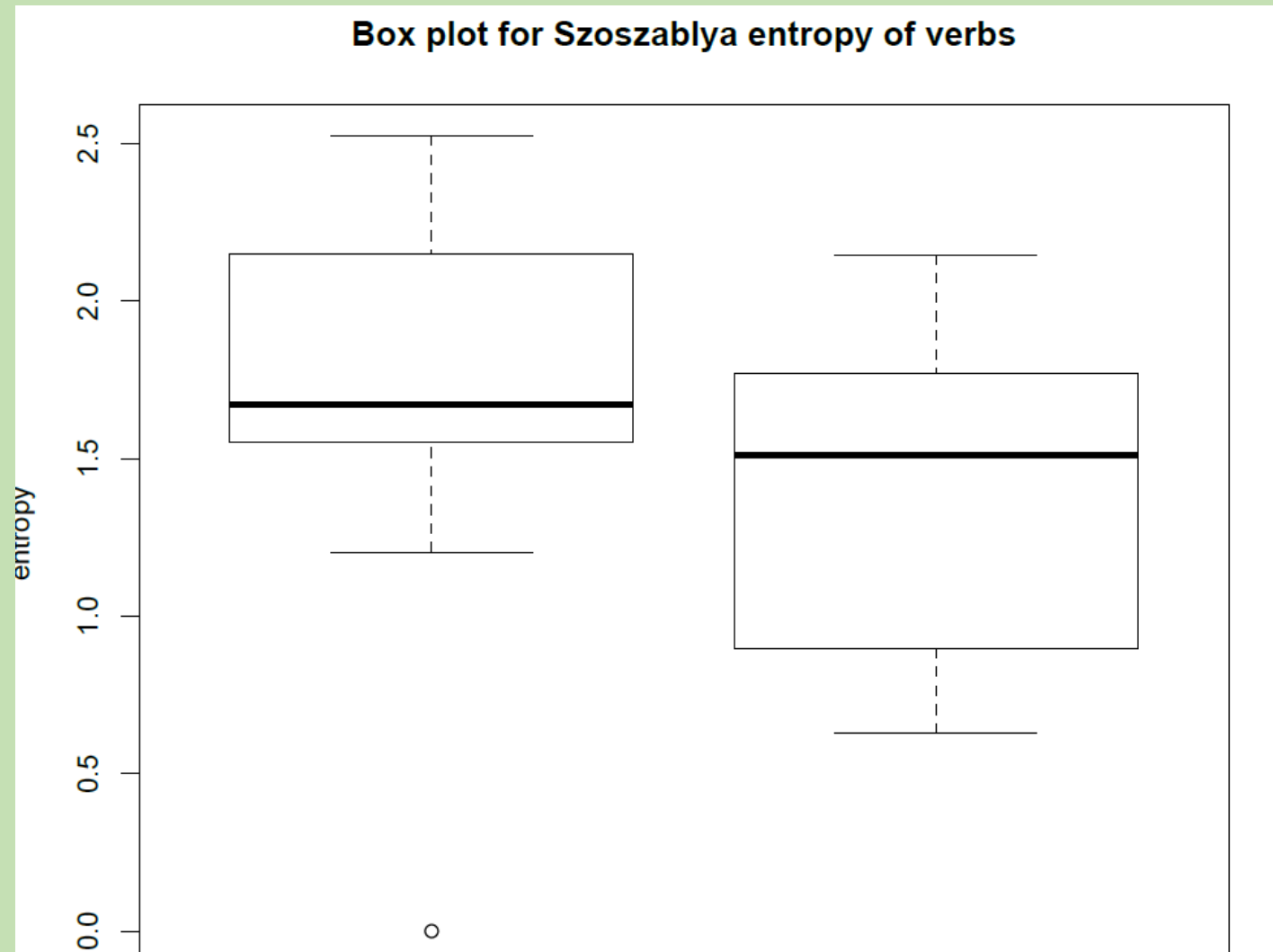
Két korpuszból elég közeli entrópia értékek



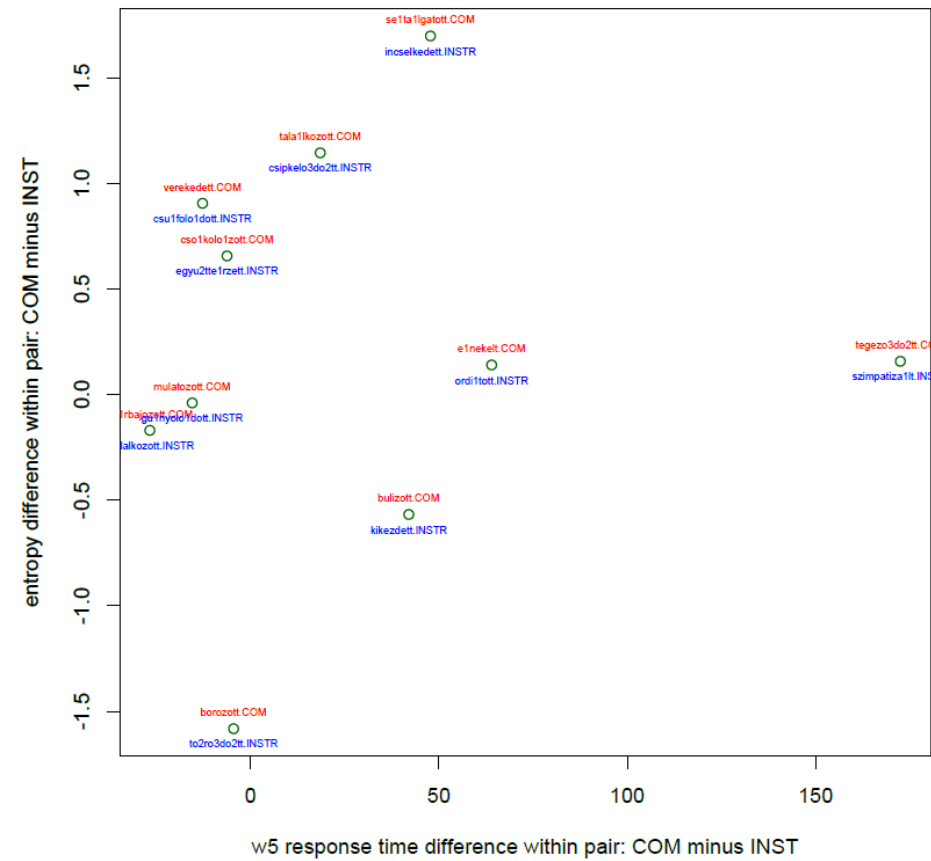
Entropy values for COM and INSTR verbs



A COMMIT entrópia magasabb sétált szimpatizált



Relationship between within verb-pair relative response time and entropy



Predikciós összefoglal ás



1. A predikciós modellek célja a jövőbeli események előrejelzése a múltbeli adatok alapján.



2. A predikciós modellek általában a múltbeli adatokból tanulnak, és a jövőbeli adatok alapján tesztelik a modellet.



3. A predikciós modellek általában a múltbeli adatokból tanulnak, és a jövőbeli adatok alapján tesztelik a modellet.

4. A predikciós modellek általában a múltbeli adatokból tanulnak, és a jövőbeli adatok alapján tesztelik a modellet.

Értelmezés

Az entrópia fontos mozzanat a nyelvi folyamatoknak.

Tanulási szempontból a szótári entrópia hatások azt sugallják, hogy mintavételes statisztikai eljárásokkal a környezeti szókincs leképeződik az egyéni mentális szótárban.

Feldolgozási tekintetben a „megértés” egyik mozzanat az a pont, amikor lecsökken a bizonytalanság. Ez igaz szavak, szóalakok és mondatok viszonyában is.

Akkor a
megértés út
az
unalomhoz?

A bizonytalanság redukcióval párhuzamosan a referenciák és tényállások strukturális alapú feldolgozása a való világ bizonytalanságainak felel meg. A feldolgozás 'belső' oldala a bizonytalanság csökkentését keresi, míg a 'külső oldala' az újdonsággal telei világra nyitott.

Remélem ez csökkentett bizonyos entrópiát

De új bizonytqlanságokat hozott
létre

